Correspondence article

Functional Analysis of All Salmonid Genomes (FAASG): an international initiative supporting future salmonid research, conservation and aquaculture

The FAASG Consortium: Daniel J. Macqueen ^ [1], Craig R. Primmer ^ [2], Ross D. Houston ^ [3], Barbara F. Nowak ^ [4], Louis Bernatchez [5], Steinar Bergseth [6], William S. Davidson * [7], Cristian Gallardo-Escárate [8], Tom Goldammer [9], Yann Guiguen [10], Patricia Iturra [11], James W. Kijas [12], Ben F. Koop * [13], Sigbjørn Lien * [14], Alejandro Maass [15,16], Samuel A.M. Martin [1], Philip McGinnity [17], Martin Montecino [18,19], Kerry A. Naish [20], Krista M. Nichols [21], Kristinn Ólafsson [22], Stig W. Omholt [14, 23], Yniv Palti [24], Graham S. Plastow [25], Caird E. Rexroad 3rd [26], Matthew L. Rise [27], Rachael J. Ritchie [28] Simen R. Sandve [14], Patricia M. Schulte [29], Alfredo Tello [30], Rodrigo Vidal [31], Jon Olav Vik [14], Anna Wargelius [32], José Manuel Yáñez [33]

[1] *Institute of Biological and Environmental Sciences, University of Aberdeen, Aberdeen AB24 2TZ, United Kingdom.*
[2] *Department of Biology, University of Turku, 20014, Finland.*
[3] *The Roslin Institute and Royal (Dick) School of Veterinary Studies, The University of Edinburgh, Midlothian, EH25 9RG, UK.*
[4] *Institute for Marine and Antarctic Studies, University of Tasmania Launceston, Tasmania, Australia.*
[5] *Département de biologie, Institut de Biologie Intégrative et des Systèmes (IBIS), Université Laval, Québec, Canada, G1V 0A6.*
[6] *The Research Council of Norway, Drammensveien 288, P.O.Box 564, NO-1327 Lysaker, Norway.*
[7] *Department of Molecular Biology and Biochemistry, Simon Fraser University, Burnaby, British Columbia V5A 1S6, Canada.*
[8] *Laboratory of Biotechnology and Aquatic Genomics, Interdisciplinary Center for Aquaculture Research, Department of Oceanography, Universidad de Concepción, Concepción 4030000, Chile*
[9] *Leibniz Institute for Farm Animal Biology, Institute for Genome Biology, Fish Genetics Unit, Wilhelm-Stahl-Allee 2, 18196, Dummerstorf, Germany.*
[10] *INRA, UR1037 Fish Physiology and Genomics, Rennes, France.*
[11] *Faculty of Medicine, University of Chile, Santiago 8380453, Chile.*
[12] *CSIRO Agriculture, St Lucia, QLD, 4067, Australia.*
[13] *Department of Biology, University of Victoria, Victoria, British Columbia V8W 3N5, Canada.*
[14] *Centre for Integrative Genetics (CIGENE), Department of Animal and Aquacultural Sciences, Norwegian University of Life Sciences, Ås NO-1432, Norway.*
[15] *Center for Mathematical Modelling, Department of Mathematical Engineering, University of Chile, Santiago 8370456, Chile.*
[16] *Center for Genome Regulation, University of Chile, Santiago 8370456, Chile.*
[17] *School of Biological, Earth and Environmental Sciences, University College Cork, Cork, Ireland.*
[18] *Center for Biomedical Research, Universidad Andres Bello, Santiago 8370146, Chile*
[19] *FONDAP Center for Genome Regulation, Faculty of Biological Sciences and Faculty of Medicine, Universidad Andres Bello, Santiago 8370146.*
[20] *School of Aquatic and Fishery Sciences, University of Washington, Box 355020, Seattle, WA 98195, USA.*
[21] *Conservation Biology Division, Northwest Fisheries Science Center, National Marine Fisheries Service, National Oceanic and Atmospheric Administration, 2725 Montlake Blvd E, Seattle, WA, 98112, USA.*
[22] *Matis Ltd, Vínlandsleið 12, 113 Reykjavík, Iceland.*
[23] *NTNU - Norwegian University of Science and Technology, NO-7491 Trondheim, Norway.*
[24] *National Center for Cool and Cold Water Aquaculture, USDA ARS, 11861 Leetown Road, Kearneysville, WV, 25430, USA.*
[25] *Department of Agricultural, Food, and Nutritional Science, University of Alberta, Edmonton, AB, Canada.*
[26] *Office of National Programs, USDA ARS, 5601 Sunnyside Avenue, Beltsville, Maryland 20705-5148.*
[27] *Department of Ocean Sciences, Memorial University of Newfoundland, 1 Marine Lab Road, St. John's, NL, A1C 5S7, Canada.*
[28] *Genome British Columbia, Suite 400 – 575, West 8th Avenue, Vancouver BC V5Z 0C4, Canada.*
[29] *Department of Zoology, University of British Columbia, 6270 University Blvd., Vancouver, BC, V6T 1Z4, Canada.*
[30] *Instituto Tecnológico del Salmón S.A., INTESAL de SalmonChile, Chile.*
[31] *Laboratory of Molecular Ecology, Genomics, and Evolutionary Studies, Department of Biology, University of Santiago, Santiago 9170022, Chile.*
[32] *Institute of Marine Research, P.O. Box 1870, Nordnes, NO-5817 Bergen, Norway.*
[33] *Faculty of Veterinary and Animal Sciences, University of Chile, Av. Santa Rosa 11735, Santiago, Chile & Aquainnovo, Cardonal s/n, Puerto Montt, Chile.*

^ Authors for correspondence on this article:

Craig R. Primmer (craig.primmer@utu.fi)
Daniel J. Macqueen (daniel.macqueen@abdn.ac.uk)
Ross D. Houston (ross.houston@roslin.ed.ac.uk)
Barbara F. Nowak (b.nowak@utas.edu.au)

* FAASG corresponding scientists:

Willie S. Davidson (Address as given above; email: william_davidson@sfu.ca)
Sigbjørn Lien (Address as given above; email: sigbjorn.lien@nmbu.no)
Ben F. Koop (Address as given above; email: bkoop@uvic.ca)

Abstract

We describe an emerging initiative - the 'Functional Analysis of All Salmonid Genomes' (FAASG), which will leverage the extensive trait diversity that has evolved since a whole genome duplication event in the salmonid ancestor, to develop an integrative understanding of the functional genomic basis of phenotypic variation. The outcomes of FAASG will have diverse applications, ranging from improved understanding of genome evolution, through to improving the efficiency and sustainability of aquaculture production, supporting the future of fundamental and applied research in an iconic fish lineage of major societal importance.

Keywords: Salmonid fish; Genome function; Comparative approach; Phenotype & data standardization; Evolution & Ecology; Physiology; Aquaculture; Whole genome duplication; Data sharing; Open access

1. The importance of salmonid fish: from evolution to sustainable food production

Salmonids have combined scientific, societal and economic importance that is unique among fish (reviewed in [1]). They are naturally distributed in fresh and marine habitats throughout the Northern hemisphere and have been introduced to South America, Australia, Africa and the Middle East. They perform key ecological functions, e.g. [2], but many populations are declining, and extensive effort is being directed towards their conservation and management, especially with respect to anthropogenic-driven change, e.g. [3]. Salmonids include at least 70 species (but are sometimes classified as > 200), possessing a rich diversity of adaptations and life-history strategies [4]. The great phenotypic diversity amongst salmonids provides an excellent study system to understand adaptive divergence and ecological speciation [4, 5] and was potentially facilitated by a whole genome duplication (WGD) in their common ancestor ~95 Mya [6, 7]. Salmonid aquaculture and capture fisheries (mainly of Atlantic salmon and *Oncorhynchus* species) play an important role in the economic and/or food security of several nations, accounting for 7.2 / 16.6% of all traded fish in terms of share by weight / value [8].

Macqueen et al. *Functional Analysis of All Salmonid Genomes (FAASG)*

2. Rationale for the FAASG initiative

The FAASG initiative follows the recent publication of the genomes of Atlantic salmon (*S. salar*) [9] and rainbow trout (*O. mykiss*) [10], which have proved invaluable to salmonid researchers (section 3) and establish a solid foundation for generating reference genome sequences for other salmonid species (Fig. 1). The next step for salmonid research is to annotate genome function, considering species and populations of major scientific interest (sections 6, 7). This will lay foundations to understand how genotypes are translated to phenotypes via different layers of regulation of gene and protein expression. Covering a broad diversity of research in salmonid biology will aid this action and is best achieved by involving the broadest possible research community (section 9). FAASG will follow principles established by the 'Functional Annotation of Animal Genomes' (FAANG) consortium (section 5) [11], using standardized approaches for functional annotation, including bioinformatics protocols and pipelines exploiting knowledge from other species and through an array of experimental assays (Table 1, section 7). However, the FAASG framework (section 6, Fig. 1) will also exploit unique features of salmonid biology, including recent WGD and extensive phenotypic variation at both macro- and micro-evolutionary timescales, to generate broad mechanistic insights into genome evolution and adaptation.
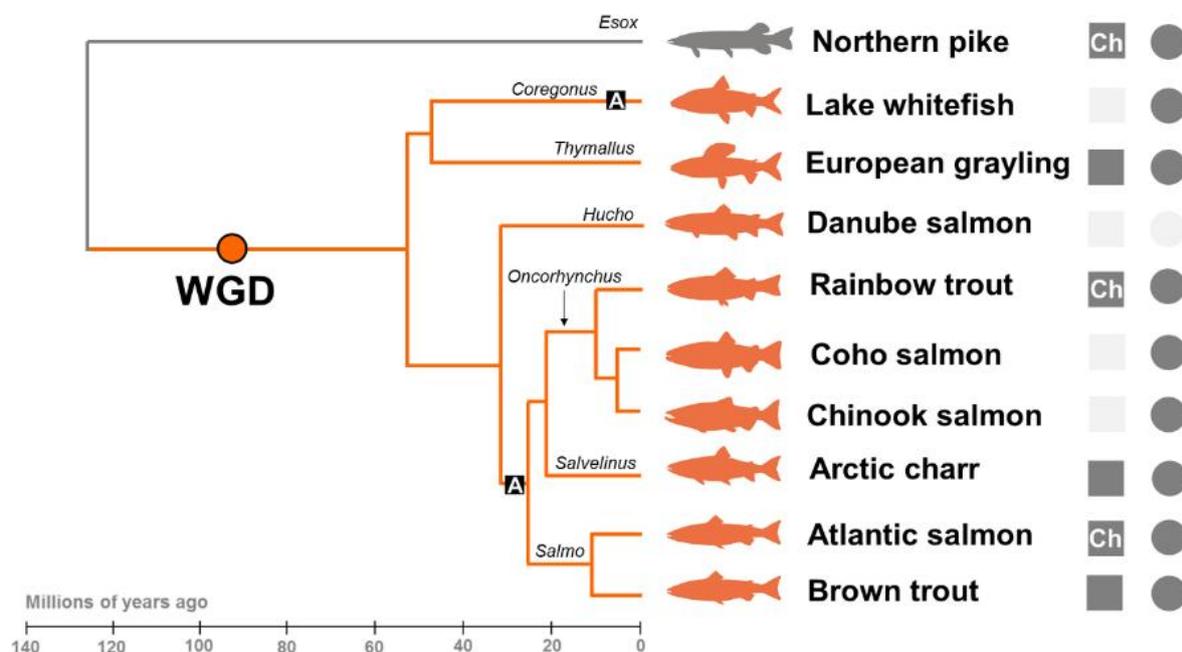

Figure 1. The comparative-evolutionary framework of FAASG. Shown are the main target lineages for functional annotation (see Table 1) and their evolutionary relationships (time-calibrated tree after [7]). The selected species come from all three salmonid subfamilies. The position of the salmonid-specific WGD is highlighted (after [7, 9-10]), along with Latin names of genera. Additional salmonid species that are future potential targets for functional annotation are not shown. Two lineages where anadromous life-history evolved are highlighted 'A' (after [47]). Relevant 'omics resources are shown to the right of the tree - finalized (dark grey) or under active development (light grey), with 'Ch' highlighting chromosome-anchored genome assemblies.

Macqueen et al. *Functional Analysis of All Salmonid Genomes (FAASG)*

Table 1. Levels of genome-wide functional annotation within the FAASG framework

Class of variation	Context	Origin of data	Goal
Genomic sequence	Phylogeny-wide	Comparative analysis	Define fixed substitutions across species including for WGD gene duplicates. Assign to different classes: exonic, intronic, regulatory, synonymous vs. non-synonymous; radical vs. conservative non-synonymous and divergent from ancestral state Identify differences in structural genomic variation among species and describe its evolution Associate sequence/structural genome variation with epigenetic, transcriptomic and proteomic variation
Genomic sequence	Population-level	Genome-resequencing	Define SNPs and structural genome variation within species. Assign to different classes: as above Associate sequence/structural genome variation with epigenetic, transcriptomic and proteomic variation
Epigenetic (DNA methylation)	Phylogeny-wide and population level	Assays described in Table S1.	Generate DNA methylome maps and define their regulation across tissues, developmental stages and common-garden physiological manipulations Associate changes in methylation with all forms of genomic, transcriptomic, proteomic and other classes of epigenetic variation
Epigenetic (histone modifications)	Phylogeny-wide and population level	Assays described in Table S1	Define a range of histone marks and their regulation across tissues, developmental stages and common-garden physiological manipulations Associate variation in histone marks with all forms of genomic, transcriptomic, proteomic and other classes of epigenetic variation
Epigenetic (chromatin biology)	Phylogeny-wide and population level	Assays described in Table S1	Generate maps of DNA accessibility and define their regulation across tissues, developmental stages and common-garden physiological manipulations Associate changes in chromatin structure with all forms of genomic, transcriptomic, proteomic and other classes of epigenetic variation
RNA expression	Phylogeny-wide and population level	RNAseq - potentially stranded protocols (see Table S1)	Define expression of miRNA, mRNA and non-coding RNA across adult tissues, developmental stages and common-garden physiological manipulations [1] Associate transcriptomic variation to all forms of genomic, epigenetic and proteomic variation
Protein level	Phylogeny-wide and population level	Various possible mass spectrometer platforms – bottom up approach	Define proteome across tissues, developmental stages and common-garden physiological manipulations Associate proteomic variation to all forms of genomic, transcriptomic and epigenetic variation

3. Genome-led science in salmonids: progress, challenges and unresolved questions

Notable progress in understanding of salmonid biology has stemmed from sequencing two salmonid genomes, as well as that of the northern pike *Esox lucius* [12], a sister lineage that did not undergo the salmonid-specific WGD (Fig. 1). Genome-wide analyses have offered key insights into the remodelling and divergence of duplicated genome content and functions during the post-WGD rediploidization process [9, 10]. Population genomics has been revolutionized by genotyping-by-sequencing, whole genome re-sequencing and high-density SNP arrays [13-15], used for example to discover SNPs near the *vgll3* gene that explain 40% of the variation in sea-age at maturity [16, 17], genomic variation explaining the timing of migration [18] and adaptive population differentiation in immune function [19]. Population genomics is now routinely applied in salmonids without a genome sequence, by exploiting conserved synteny with rainbow trout or Atlantic salmon, e.g. [20-23]. Genome-wide approaches have also been applied to improve the accuracy of selection for key production traits (e.g. disease resistance) in breeding programs, either through genomic selection [24-26] or by characterization of major effect loci, e.g. [27, 28]. Further, the salmonid and pike genomes have been used to progress understanding of salmonid phylogeny and species radiation [7] and facilitate characterization of the molecular basis and post-WGD evolution of several physiological systems, including smoltification [29], growth [30], immunity [19, 31, 32] and olfaction [33]. Finally, the recent demonstration of successful genome editing in salmonids [34-37] opens the door for validation of candidate functional genomic elements and causative polymorphisms. Genome editing also has potential to address certain challenges in aquaculture, by creating or introgressing alleles into farmed populations, and by expediting the selection for existing beneficial alleles [38].

Nonetheless, salmonid research and its applications have only just begun to exploit the possibilities of genome-led science. Undoubtedly, a number of unresolved questions and important challenges can be addressed through the FAASG initiative (Box 1).

Macqueen et al. *Functional Analysis of All Salmonid Genomes (FAASG)*

Box 1. The role of functional genome annotation in addressing key challenges for salmonid research and application. Below we list selected key questions, highlight their importance, then briefly describe (*in italics*) how the FAASG initiative will help address them.

Aquaculture

What is the functional genetic basis of key performance traits for salmonid aquaculture?

Few causative variants underlying performance trait QTL have been identified. Knowledge of the precise functional variants underpinning QTL will inform the biology of these traits, and facilitate cost-effective selection for favorable alleles.

Genome annotation is essential to prioritize candidate causal variants. Many traits are influenced by non-coding variants influencing gene expression. The FAASG initiative will aid identification and prioritization of QTL-region variants for key traits.

How can we optimize genomic selection for genetic improvement in aquaculture breeding programs?

Genomic selection can accelerate genetic gain for traits important to sustainable and profitable aquaculture, such as host resistance to infectious diseases. Predicting breeding values in distant relatives to the training population is challenging, thus necessitating frequent, expensive phenotypic tests.

The likelihood of SNPs having a functional effect on a trait can be estimated using FAASG functional annotation data. These SNPs can be prioritized in genotyping panels to enable improved prediction accuracy, and persistency of that accuracy, across diverse genetic backgrounds and multiple generations.

What is the functional genetic basis of recent domestication in salmonid species?

Salmonids are excellent models for the genomic basis of recent domestication, facilitating discovery of genetic variation of importance in adaptation to aquaculture environments. These outcomes can improve hatchery management, health and welfare of farmed fish, and have implications for interactions with wild populations.

Domestication is likely to have a polygenic basis and be largely due to modification of gene regulation and may also have some epigenetic basis. Functional annotation is essential for researchers to identify sequence and epigenomic variation linked to domestication and response to artificial selection.

How can genome editing technology contribute to improved aquaculture production?

Genome editing technology, notably CRISPR-Cas9, has potential to enhance aquaculture production directly, by creating or introducing favorable alleles into farmed populations. While regulatory and public acceptance is required, the potential is highlighted by several high profile successes in terrestrial livestock.

Choosing the correct target to edit is essential, and requires accurate annotation of the reference genome. A function of a SNP, epigenetic mark, non-coding RNA, coding RNA or whole protein can be determined using gene editing. The technology can also be applied to demonstrate causality of variants underlying QTL.

What is the long term impact of aquaculture escapees on wild populations?

Evaluating and understanding the impacts of aquaculture escapees on wild populations supports risk assessment for the use of native and non-native strains in culture.

FAASG will improve understanding of the functional differences among populations resulting from genomic variation, and will guide development of tools to effectively track and monitor the genetic impact of escapees on wild populations.

Macqueen et al. *Functional Analysis of All Salmonid Genomes (FAASG)*

How can measurement of salmonid health and welfare in aquaculture be improved?

Appropriate biomarkers of stress, health and growth status in salmonid aquaculture are currently difficult to define and far from comprehensive.

An improved understanding of the genetic and epigenetic regulation of key physiological systems supporting fish health will be guided by the annotated genomes, networks and comparative biology, and will facilitate development of tools to help monitor animal wellbeing in culture.

Ecology, evolution and physiology

What role did the whole genome duplication and subsequent rediploidization play in salmonid evolution and species radiation?

This is a long-standing question of fundamental importance to our understanding of salmonid biology and the role of WGDs in evolution more generally.

Comparative genomic annotation will improve understanding of how sequence and functional variation arising post-WGD is coupled to trait evolution, including lineage-specific evolution of anadromous life-history and species radiation.

How important is genetic vs. epigenetic variation in regulating trait variability?

Rapid phenotypic divergence and phenotypic plasticity is a hallmark of many salmonids species, yet remains poorly-characterized. An improved understanding of heritable epigenetic variation and its interaction with both genetic and environmental variation can be exploited in both conservation and aquaculture.

Functional annotation of epigenetic marks in salmonid genomes, and studies into the role of epigenetic regulation in determining trait variation and phenotypic plasticity are key goals of the FAASG initiative.

What is the genomic basis of response and adaptation to natural and anthropogenic stressors?

Human-induced environmental changes, including climate change, are already negatively affecting salmonid populations. Understanding the role of genetic and epigenetic variation in physiological response to these changes will be key to predicting, and potentially mitigating, these effects.

Improved understanding of the functional genomic basis of differential responses to environmental stressors in salmonids may be applied to inform forecasting, mitigation and remedial strategies for challenges associated with climate change

What role do 'non-coding' RNAs have in generating phenotypic variation?

The functions of non-coding RNAs are poorly understood in salmonids. The greater retention of miRNAs in comparison to duplicated genes after WGD suggests important functions in coping with a duplicated genome. Non-coding RNAs may regulate traits of interest to aquaculture and evolutionary biology.

Comparative functional annotation in salmonids will highlight the location and role of non-coding RNAs in regulation of gene expression and downstream regulation of complex traits.

How many salmonid species exist, and how can we distinguish them?

The actual number of salmonid species is unknown. Habitat-dependent phenotypes can suggest different species, but genomics and functional genomics methods are ultimately required to answer this question.

Diverse salmonid species and populations will be targeted in FAASG, providing comparative genome sequences and annotations. This will facilitate development and application of species-specific markers to assess the quantity and diversity of species in the salmonid family.

Macqueen et al. *Functional Analysis of All Salmonid Genomes (FAASG)*

4. Traits of crosscutting relevance: from aquaculture to evolution (and beyond)

Several traits of importance to aquaculture show extensive natural variation among salmonid species and populations, including disease resistance, growth rate, the control of sexual determination and maturation, and the physiological transition from fresh to saltwater. These traits have crosscutting relevance to multiple scientific fields, both fundamental and applied, and the dissection of their functional genomic architecture under the FAASG initiative will help address challenges faced by the aquaculture sector, along with long-standing research questions. Accordingly, the outcomes of FAASG will facilitate selection of aquaculture strains with improved disease resistance and higher product quality that reach market earlier [39-41], while explaining the evolutionary role of trait variation in wild populations [16, 42-43] and informing management actions influencing population resilience, conservation, and re-introduction [23, 44-46]. Comparing the outcomes of artificial vs. natural selection on functional pathways under different conditions will also help dissect the genetic architecture of traits. For example, different populations will often share genetic variation influencing a trait, but aquaculture and wild conditions impose divergent selective pressures, leading to unique, yet complementary opportunities to understand natural selection and domestication.

5. Rationale for linking with FAANG

The FAANG consortium aims to produce comprehensive maps of the functional elements in the genomes of domesticated animal species [11], building on the ENCODE project [47]. Underpinning principles of both consortia include use of robust, standardized experimental protocols based on defined tissues or cell types. These principles apply to both 'wet lab' experiments and bioinformatic analyses of data, which provides a comprehensive, reliable, open-access resource for a wide community. The FAASG initiative will link to FAANG, adhere to these principles, and utilise and build on the FAANG protocols and pipelines to avoid redundancy. FAANG is focussed on livestock species with high-quality reference genomes (chicken, pig, cattle and sheep), but with scope for inclusion of other species. The initial focus of FAASG will be the key farmed salmonids (Atlantic salmon and rainbow trout), but will expand to a broader range of lineages of interest to conservation, management and evolution (Fig. 1). In doing so, the initiative will harness wider diversity within a comparative context (section 6) to understand the evolution of functional genome elements following species radiation and WGD. FAASG will provide a FAANG-type model for other lineages with recently-developed genome assemblies, the number of which is rapidly increasing. There will also be great scope for cross-talk between FAASG and research communities for model fish species where functional annotation is advanced, including zebrafish (http://zfin.org/).

Macqueen et al. *Functional Analysis of All Salmonid Genomes (FAASG)*

6. The FAASG framework

The initial approach of FAASG will exploit the full phylogenetic framework of salmonid evolution, documenting functionally important sequence variation and data derived from a core set of experimental assays (section 7) across at nine eight salmonid species (from six out of ten true genera) and northern pike (Fig. 1) and under experimental conditions representative of the traits listed in section 4. This sampling traverses the diversification of salmonid lineages and evolutionary origins of anadromy, a life-history strategy that evolved at least twice independently [48] (Fig. 1) and facilitated species radiations [7]. FAASG will address micro-evolutionary variation by contrasting wild populations that evolved divergent phenotypes over thousands of years and aquaculture vs wild strains separated by relatively few generations (Fig. 1). This combination of experimental assays and evolutionary analyses done across the salmonid phylogeny (Section 7) will be applied to assess 'genome function', thereby addressing a potential shortcoming of the original interpretations of the ENCODE data [49].

7. Data and assays

The assays of FAASG are described in Table 1 (also, see Table S1). Annotating distinct classes of sequence variation will identify the genome-wide evolution of orthologous protein-coding genes, along with the large number of retained functional gene duplicates (> 50% of those created) from WGD [9-10]. Population-level sequence variation will inform the role of functional elements in recent phenotypic divergence and adaptation (Table 1). The inclusion of Northern pike (Fig. 1) will enable the ancestral (non-duplicated) state of sequence variation to be inferred, including the direction of divergence between duplicated genes. Comparisons to more distantly related fish with well-annotated genomes, including zebrafish *Danio rerio* [50], three-spined stickleback *Gasterosteus aculeatus* [51], spotted gar *Lepisosteus oculatus* [52], European seabass *Dicentrarchus labrax* [53], and Asian seabass *Lates calcarifer* [54], will allow salmonid-specific changes to be contextualized in the broader framework of teleost evolution, especially with respect to the earlier WGD event in the teleost ancestor.

Transcriptome and proteome phenotypes will be characterized for a panel of tissues and developmental stages, sampled from both sexes under common garden conditions using standardized sampling and analytical protocols that distinguish divergence in expression of duplicated loci [9-10] (e.g. RNA extraction, quality control, library preparation, choice of sequencing platform, and bioinformatic analyses). Discerning the regulation and evolution of transcript complexity (e.g. non-coding, miRNome and splice variants) will necessitate stranded approaches [55] and may be facilitated by capture of full-length transcripts through single

Macqueen et al. *Functional Analysis of All Salmonid Genomes (FAASG)*

molecule real-time sequencing [56]. Standardized proteome expression profiling will also be performed after experimental separation of different cellular fractions.

FAASG will implement genome-wide experimental assays being used or considered under FAANG [11] (Table 1, S1), including: 1) methylation at nucleotide-level resolution (several approaches available, e.g. [57-58]), 2) chromosome accessibility and architecture (via ATAC-Seq [59]), DNase I footprinting [60], or ChIP-seq approaches), 3) histone modifications (using ChIP-seq approaches [61-62]), 4) genome conformation (via Hi-C [63]) and 5) transcription factor binding occupancy (via ChIP-seq approaches [64]). Many of these assays have yet to be employed in salmonids (Table S1). However, several studies have laid the groundwork for such efforts, and no technical limitations are expected given that these approaches rely on generic techniques and represent conserved features of molecular biology. Initial experiments in Atlantic salmon and rainbow trout will be conducted in the context of regulation across tissues and developmental stages. Assays incorporating different lineages, populations, and physiological manipulations will follow within the wider comparative-phylogenetic framework. Targeted genome editing can subsequently be used to infer causality of sequence variants and functional genomic elements.

8. Importance of standardized phenotypic data

Informative genome functional annotation will necessitate standardized measurement and recording of both ecologically and production-relevant traits (section 4) and for the effects of plasticity [65] to be controlled. Comparisons of the genetic architectures for complex phenotypes are confounded not only by the environment in which traits are measured, but also by how those traits are quantified. We view common garden experiments, performed under agreed standardized conditions and treatments, as central to the collection of high-quality phenotype data. Salmonids are well-suited for common-garden experiments as they possess external fertilization, high fecundity, and have high survival rates in captivity. In addition, facilities are widely available to raise large numbers of fish under a range of controlled experimental contexts. Such features also facilitate robust and powerful studies to dissect the quantitative genetic basis of complex traits. The standardized recording of both ecologically and production-relevant phenotypes and cataloguing of functional and phenotypic responses, e.g. within the Gene Ontology framework is also high priority. Standardised phenotypic assays will also help interpret the molecular basis of phenotypic variation observed in the numerous wild populations gained by long-term data series, e.g. [66].

Macqueen et al. *Functional Analysis of All Salmonid Genomes (FAASG)*

9. Engaging the research community

The FAASG initiative will promote inclusiveness among all stakeholders and draw in expertise in aquaculture, bioinformatics/biostatistics, genetics, molecular biology, functional genomics, physiology, ecology and conservation, ensuring quality at all levels. A website (http://www.faasg.org/) will report progress, including experimental and computational protocols, publications and datasets, along with contact information for researchers or funders wishing to contribute to efforts. The initiative is being advertised at several scientific conferences (see http://www.faasg.org/) to promote wider awareness.

Declarations:

- List of abbreviations

FAANG- Functional Annotation of Animal Genomes

FAASG- Functional Annotation of All Salmonid Genomes

WGD- Whole genome duplication

- Ethics (and consent to participate): Not applicable

- Consent to publish: Not applicable

- Competing interests

SB and RJR are respective employees of the Research Council of Norway and Genome British Columbia and made contributions to the scientific writing of this article and the general scientific development of the FAASG initiative.

- Funding

The first FAASG meeting (June 2016, http://www.faasg.org/publications/toronto-workshop-report/) was organized and supported by the International Cooperation to Sequence the Atlantic Salmon Genome (ICSASG). ICSASG funders: The Research Council of Norway (RCN), The Norwegian Seafood Research Fund (FHF), Genome British Columbia (Canada) GBC), The Chilean Economic Development Agency (CORFO) and the Innova Chile Committee (InnovaChile). SB and RJR are respective employees of the Research Council of Norway and Genome British Columbia and made contributions to the scientific writing of this article and the general scientific development of the FAASG initiative.

- Authors' contributions

Coordinated and drafted the manuscript: DJM, CRP, RDH, BFN. SRS contributed to Figure 1. All authors contributed to manuscript writing and read and approved the final version of this article.

- Availability of data and materials: not applicable

Macqueen et al. *Functional Analysis of All Salmonid Genomes (FAASG)*

References

1. Davidson WS, Koop BF, Jones SJ, Iturra P, Vidal R, Maass A, Jonassen I, Lien S, Omholt SW. Sequencing the genome of the Atlantic salmon (*Salmo salar*). Genome Biol. 2010;11(9):403.
2. Schindler DE, Scheuerell MD, Moore JW, Gende SM, Francis TB, Palen WJ. Pacific salmon and the ecology of coastal ecosystems. Front Ecol Environ 2003;1(1):31-7.
3. Waples RS, Hendry AP. Special Issue: Evolutionary perspectives on salmonid conservation and management. Evol Appl. 2008;1(2):183-8.
4. Hendry AP, Stearns SC. Evolution Illuminated: Salmon and Their Relatives. New York: Oxford University Press; 2004.
5. Bernatchez L, Renaut S, Whiteley AR, Campbell D, Derôme N, Jeukens J, et al. On the origin of species: Insights from the ecological genomics of whitefish. Philos Trans R Soc Lond B Biol Sci. 2010 365(1547):1783-800.
6. Allendorf FW, Thorgaard GH. Tetraploidy and the evolution of salmonid fishes. In: Turner BJ, editor. Evolutionary genetics of fishes. New York: Plenum Press; 1984. p. 1–53.
7. Macqueen DJ, Johnston IA. A well-constrained estimate for the timing of the salmonid whole genome duplication reveals major decoupling from species diversification. Proc Biol Sci. 2014;281(1778):20132881.
8. FAO. 2016. The State of World Fisheries and Aquaculture 2016. Contributing to food security and nutrition for all. Rome. 200 pp.
9. 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(7602):200-5.
10. Berthelot C, Brunet F, Chalopin D, Juanchich A, Bernard M, Noël B, et al. The rainbow trout genome provides novel insights into evolution after whole-genome duplication in vertebrates. Nat Commun. 2014;5:3657.
11. Andersson L, Archibald AL, Bottema CD, Brauning R, Burgess SC, Burt DW, et al. Coordinated international action to accelerate genome-to-phenome with FAANG, the Functional Annotation of Animal Genomes project. Genome Biol. 2015;16:57.
12. Rondeau EB, Minkley DR, Leong JS, Messmer AM, Jantzen JR, von Schalburg KR, et al. The genome and linkage map of the northern pike (*Esox lucius*): conserved synteny revealed between the salmonid sister group and the Neoteleostei. PLoS One. 2014;9(7):e102089.
13. Houston RD, Taggart JB, Cézard T, Bekaert M, Lowe NR, Downing A, et al. Development and validation of a high density SNP genotyping array for Atlantic salmon (*Salmo salar*). BMC Genomics. 2014;15:90.
14. Palti Y, Gao G, Liu S, Kent MP, Lien S, Miller MR, Rexroad CE 3rd, et al. The development and characterization of a 57K single nucleotide polymorphism array for rainbow trout. Mol Ecol Resour. 2015;15(3):662-72.
15. Yáñez JM, Naswa S, López ME, Bassini L, Correa K, Gilbey J, et al. Genomewide single nucleotide polymorphism discovery in Atlantic salmon (*Salmo salar*): validation in wild and farmed American and European populations. Mol Ecol Resour. 2016;16(4):1002-11.

16. Barson NJ, Aykanat T, Hindar K, Baranski M, Bolstad GH, Fiske P, et al. Sex-dependent dominance at a single locus maintains variation in age at maturity in salmon. Nature. 2015;528(7582):405-8.

17. Ayllon F, Kjærner-Semb E, Furmanek T, Wennevik V, Solberg MF, Dahle G, et al. The vgll3 locus controls age at maturity in wild and domesticated Atlantic salmon (*Salmo salar* L.) males. PLoS Genet. 2015;11(11):e1005628.

18. Hess JE, Zendt JS, Matala AR, Narum SR. Genetic basis of adult migration timing in anadromous steelhead discovered through multivariate association testing. Proc Biol Sci. 2016;283(1830).

19. Kjærner-Semb E, Ayllon F, Furmanek T, Wennevik V, Dahle G, Niemelä E, et al. Atlantic salmon populations reveal adaptive divergence of immune related genes - a duplicated genome under selection. BMC Genomics. 2016;17(1):610.

20. Seeb LW, Waples RK, Limborg MT, Warheit KI, Pascal CE, Seeb JE. Parallel signatures of selection in temporally isolated lineages of pink salmon. Mol Ecol. 2014;23(10):2473-85

21. Hecht BC, Matala AP, Hess JE, Narum SR. Environmental adaptation in Chinook salmon (*Oncorhynchus tshawytscha*) throughout their North American range. Mol Ecol. 2015;24(22):5573-95.

22. Brieuc MS, Ono K, Drinan DP, Naish KA. Integration of Random Forest with population-based outlier analyses provides insight on the genomic basis and evolution of run timing in Chinook salmon (*Oncorhynchus tshawytscha*). Mol Ecol. 2015;24(11):2729-46.

23. Nichols KM, Kozfkay C, Narum S. Genomic signatures of migration in *Oncorhynchus nerka* to inform conservation and management of endangered stocks. Evol Appl. 2016. *In press.*

24. Odegård J, Moen T, Santi N, Korsvoll SA, Kjøglum S, Meuwissen TH. Genomic prediction in an admixed population of Atlantic salmon (*Salmo salar*). Front Genet. 2014;5:402.

25. Tsai HY, Hamilton A, Tinch AE, Guy DR, Bron JE, et al. 2016. Genomic prediction of host resistance to sea lice in farmed Atlantic salmon populations. Genet Sel Evol. 2016;48(1):47.

26. Vallejo RL, Leeds TD, Fragomeni BO, Gao G, Hernandez AG, Misztal I, et al. Evaluation of genome-enabled selection for bacterial cold water disease resistance using progeny performance data in rainbow trout: insights on genotyping methods and genomic prediction models. Front Genet. 2016;7:96.

27. 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(5):405-14.

28. Moen T, Torgersen J, Santi N, Davidson WS, Baranski M, Ødegård J, et al. Epithelial cadherin determines resistance to infectious pancreatic necrosis virus in Atlantic salmon. Genetics. 2015;200(4):1313-26.

29. Lorgen M, Casadei E, Król E, Douglas A, Birnie MJ, Ebbesson LO, et al. Functional divergence of type 2 deiodinase paralogs in the Atlantic salmon. Curr Biol. 2015;25:936-41.

30. Macqueen DJ, Garcia de la Serrana D, Johnston IA. Evolution of ancient functions in the vertebrate insulin-like growth factor system uncovered by study of duplicated salmonid fish genomes. Mol Biol Evol. 2013;30(5):1060-76.

31. Grimholt U, Hauge H, Hauge AG, Leong J, Koop BF. Chemokine receptors in Atlantic salmon. Dev Comp Immunol. 2015;49(1):79-95.

32. Wang T, Johansson P, Abós B, Holt A, Tafalla C, Jiang Y, et al. First in-depth analysis of the novel Th2-type cytokines in salmonid fish reveals distinct patterns of expression and modulation but overlapping bioactivities. Oncotarget. 2016;7(10):10917-46.

33. Tessarolo JA, Tabesh MJ, Nesbitt M, Davidson WS. 2014. Genomic organization and evolution of the trace amine-associated receptor (TAAR) repertoire in Atlantic salmon (*Salmo salar*). G3 (Bethesda). 2014;4(6):1135-41.

Macqueen et al. *Functional Analysis of All Salmonid Genomes (FAASG)*

34. Yano A, Guyomard R, Nicol B, Jouanno E, Quillet E, Klopp C, et al. An immune-related gene evolved into the master sex-determining gene in rainbow trout, *Oncorhynchus mykiss*. Curr Biol. 2012;22(15):1423-8.

35. Edvardsen RB, Leininger S, Kleppe L, Skaftnesmo KO, Wargelius A. Targeted mutagenesis in Atlantic salmon (*Salmo salar* L.) using the CRISPR/Cas9 system induces complete knockout individuals in the F0 generation. PLoS One. 2014;9(9):e108622.

36. Wargelius A, Leininger S, Skaftnesmo KO, Kleppe L, Andersson E, Taranger GL, et al. Dnd knockout ablates germ cells and demonstrates germ cell independent sex differentiation in Atlantic salmon. Sci Rep. 2016;6:21284

37. Dehler CE, Boudinot P, Martin SA, Collet B. Development of an efficient genome editing method by CRISPR/Cas9 in a fish cell line. Mar Biotechnol (NY). 2016;18(4):449-52.

38. Myers JM, Heggelund PO, Hudson G, Iwamoto RN. Genetics and broodstock management of coho salmon. Aquaculture. 2001;197:43-62.

39. Jenko J, Gorjanc G, Cleveland MA, Varshney RK, Whitelaw CB, Woolliams JA, et al. Potential of promotion of alleles by genome editing to improve quantitative traits in livestock breeding programs. Genet Sel Evol. 2015;47:55.

40. Gutierrez AP, Lubieniecki KP, Fukui S, Withler RE, Swift B, Davidson WS. Detection of quantitative trait loci (QTL) related to grilsing and late sexual maturation in Atlantic salmon (*Salmo salar*). Mar Biotechnol (NY). 2014;16(1):103-10

41. Gutierrez AP, Yáñez JM, Fukui S, Swift B, Davidson WS. 2015. Genome-wide association study (GWAS) for growth rate and age at sexual maturation in Atlantic salmon (*Salmo salar*). PLoS One. 2015;10(3):e0119730.

42. Schindler DE, Hilborn R, Chasco B, Boatright CP, Quinn TP, Rogers LA, Webster MS. Population diversity and the portfolio effect in an exploited species. Nature. 2010;465(7298):609-12.

43. Hale MC, Thrower FP, Berntson EA, Miller MR, Nichols KM. Evaluating adaptive divergence between migratory and nonmigratory ecotypes of a salmonid fish, *Oncorhynchus mykiss*. G3 (Bethesda). 2013;3(8):1273-85.

44. McGinnity P, Prodohl P, Ferguson K, Hynes R, O' Maoileidigh N, Baker N, et al. Fitness reduction and potential extinction of wild populations of Atlantic salmon, *Salmo salar*, as a result of interactions with escaped farm salmon. Proc Biol Sci. 2003;270(1532):2443-50.

45. Muhlfeld CC, Kovach RP, Jones LA, Al-Chokhachy R, Boyer MC, Leary RF, et al. Invasive hybridization in a threatened species is accelerated by climate change. Nat Clim Change. 2014;4(7):620-624.

46. Waters CD, Hard JJ, Brieuc MS, Fast DE, Warheit KI, Waples RS, et al. Effectiveness of managed gene flow in reducing genetic divergence associated with captive breeding. Evol Appl. 2015;8(10):956-71.

47. ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. Nature. 2012;489(7414):57-74.

48. Alexandrou MA, Swartz BA, Matzke NJ, Oakley TH. Genome duplication and multiple evolutionary origins of complex migratory behavior in Salmonidae. Mol Phylogenet Evol. 2013;69(3):514-2.

49. Graur D, Zheng Y, Price N, Azevedo RB, Zufall RA, Elhaik E. On the immortality of television sets: "function" in the human genome according to the evolution-free gospel of ENCODE. Genome Biol Evol. 2013;5(3):578-90.

50. Howe K, Clark MD, Torroja CF, Torrance J, Berthelot C, Muffato M, et al. The zebrafish reference genome sequence and its relationship to the human genome. Nature. 2013;496(7446):498-503.

51. Jones FC, Grabherr MG, Chan YF, Russell P, Mauceli E, Johnson J, et al. The genomic basis of adaptive evolution in threespine sticklebacks. Nature. 2012;484(7392):55-61.

52. Braasch I, Gehrke AR, Smith JJ, Kawasaki K, Manousaki T, Pasquier J, et al. 2016. The spotted gar genome illuminates vertebrate evolution and facilitates human-teleost comparisons. Nat Genet. 2016;48(4):427-37.

53. Tine M, Kuhl H, Gagnaire PA, Louro B, et al. European sea bass genome and its variation provide insights into adaptation to euryhalinity and speciation. Nat Commun. 2014;5:5770.

54. Vij S, Kuhl H, Kuznetsova IS, Komissarov A, Yurchenko AA, Van Heusden P, et al. Chromosomal-level assembly of the Asian seabass genome using long sequence reads and multi-layered scaffolding. PLoS Genet. 2016;12: e1005954.

55. Parkhomchuk D, Borodina T, Amstislavskiy V, Banaru M, Hallen L, Krobitsch S, et al. Transcriptome analysis by strand-specific sequencing of complementary DNA. Nucleic Acids Res. 2009;37(18):e123.

56. Vilfan ID, Tsai YC, Clark TA, Wegener J, Dai Q, Yi C, et al. Analysis of RNA base modification and structural rearrangement by single-molecule real-time detection of reverse transcription. J Nanobiotechnology. 2013;11:8

57. Meissner A, Gnirke A, Bell GW, Ramsahoye B, Lander ES, Jaenisch R. Reduced representation bisulfite sequencing for comparative high-resolution DNA methylation analysis. Nucleic Acids Res. 2005;33(18):5868-77.

58. Maunakea AK, Nagarajan RP, Bilenky M, Ballinger TJ, D'Souza C, Fouse SD, et al. Conserved role of intragenic DNA methylation in regulating alternative promoters. Nature. 2010;466(7303):253-7.

59. Buenrostro JD, Giresi PG, Zaba LC, Chang HY, Greenleaf WJ. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. Nat Methods. 2013;10(12):1213-8.

60. Thurman RE, Rynes E, Humbert R, Vierstra J, Maurano MT, Haugen E, et al. The accessible chromatin landscape of the human genome. Nature. 2012;489(7414):75-82.

61. Ho JWK, Jung YL, Liu T, Alver BH, Lee S, Ikegami K, et al. Comparative analysis of metazoan chromatin organization. Nature. 2014;512(7515):449-52.

62. Schmidl C, Rendeiro AF, Sheffield NC, Bock C. ChIPmentation: fast, robust, low-input ChIP-seq for histones and transcription factors. Nat Methods. 2015;12(10):963-5.

63. Rao SSP, Huntley MH, Durand NC, Stamenova EK, Bochkov ID, Robinson JT, et al. A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. Cell. 2014;159(7):1665-80.

64. Landt SG, Marinov GK, Kundaje A, Kheradpour P, Pauli F, Batzoglou S, et al. ChIP-seq guidelines and practices of the ENCODE and modENCODE consortia. Genome Res. 2012;22(9):1813-31.

65. de Villemereuil P, Gaggiotti OE, Mouterde M, Till-Bottraud I. Heredity (Edinb). Common garden experiments in the genomic era: new perspectives and opportunities. 2016;116(3):249-54.

66. McGinnity P, Jennings E, DeEyto E, Allott N, Samuelsson P, Rogan G, et al. Impact of naturally spawning captive-bred Atlantic salmon on wild populations: depressed recruitment and increased risk of climate-mediated extinction. Proc Biol Sci. 2009;276(1673):3601-10.

Macqueen et al. *Functional Analysis of All Salmonid Genomes (FAASG)*

Table S1. **Review of FAANG core/additional assays (as defined in [11]) and relevant work published in salmonids**

FAANG core assays	Assay target	Example related assays used in salmonids; assay target
Transcribed loci		
RNA-sequencing, stranded protocols [55, 70]	Transcriptome, strand polarity retained	RNA-seq, stranded protocol; transcriptome [71] RNA-seq; double stranded; transcriptome [67-69] RNA-seq; double stranded, miRNA [72] RNA-seq; double stranded; Long non-coding RNA [73] RNA-seq; double stranded; Large intergenic non-coding RNAs [74]
Chromatin accessibility and architecture		
Assay for transposase-accessible chromatin sequencing (ATAC-seq) [[59, 75]	Regions of open chromatin, localization of nucleosomes in regulatory sites and positions of DNA-binding proteins	No published examples
DNaseI footprinting [60]	Open chromatin, delineate genomic regulatory compartments	No published examples
Chromatin immunoprecipitation sequencing (ChiP-seq)	Proteins linking genome architecture to function (FAANG- highly conserved insulator-binding factor, CTCF) [74]	No published examples
Histone modification marks		
Chromatin immunoprecipitation sequencing (ChiP-seq) to detect modified histones and characterize associated sequences [61]	Histone H3 lysine 4 trimethylation (H3K4me3), identifies active gene promotors and is enriched at transcription start sites Histone H3 lysine 27 trimethylation (H3K27me3), marks genes that have been facultatively repressed through regional modification Histone H3 lysine 27 acetylation (H3K27ac), marks active regulatory elements, may discriminate active from inactive enhancers and promoters	Chromosome immune precipitation; relationship between modified histones and gene expression [77]

Macqueen et al. *Functional Analysis of All Salmonid Genomes (FAASG)*

	H3 lysine 4 monomethylation (H3K4me1), marks enhancers and other distal elements, and is enriched downstream of transcription start sites	
Additional FAANG Assays	**Assay target**	**Related assay used in salmonids; assay target**
DNA methylation, genome-wide analysis of 5-methylcytosines, nucleotide level resolution [57-58]	Epigenetic mark and regulator of gene expression	Bisulphite sequencing, nucleotide-level resolution [78]
ChiP-seq assays [64] for sequences bound by specific proteins	Transcription factor binding sites	Chromosome Immune Precipitation Assay, regulation of transcription [79-82]
Genome conformation; Hi-C [63, 83] for chromosomal conformation capture	Identify distal chromatin elements that are brought together through 3D chromosomal folding	No published examples

Additional supporting references

67. Koop BF, von Schalburg KR, Leong J, Walker N, Lieph R, Cooper GA et al. A salmonid EST genomic study: genes, duplications, phylogeny and microarrays. BMC Genomics. 2008, 9(1):1-16.

68. Pasquier J, Cabau C, Nguyen T, Jouanno E, Severac D, Braasch I et al. Gene evolution and gene expression after whole genome duplication in fish: the PhyloFish database. BMC Genomics 2016, 17(1):1-10.

69. Salem M, Paneru B, Al-Tobasei R, Abdouni F, Thorgaard GH, Rexroad CE et al. Transcriptome assembly, gene annotation and tissue gene expression atlas of the rainbow trout. PLoS ONE. 2015, 10(3):e0121778.

70. Mills JD, Kawahara Y, Janitz M. Strand-specific RNA-seq provides greater resolution of transcriptome profiling. Curr Genomics. 2013, 14(3):173-81.

71. Narum SR, Campbell NR. Transcriptomic response to heat stress among ecologically divergent populations of redband trout. BMC Genomics 2015, 16(1):1-12.

72. Andreassen R, Worren MM, Høyheim B. Discovery and characterization of miRNA genes in Atlantic salmon (*Salmo salar*) by use of a deep sequencing approach. BMC Genomics. 2013, 14:482-82.

73. Al-Tobasei R, Paneru B, Salem M. Genome-wide discovery of long non-coding RNAs in rainbow trout. PLoS One. 2016;11(2):e0148940.

74. Wang J, Fu LY, Koganti PP, Wang L, Hand JM, Ma H, Yao JB. Identification and functional prediction of large intergenic noncoding RNAs (lincRNAs) in rainbow trout (*Oncorhynchus mykiss*). Mar Biotechnol. 2016, 18(2):271-282.

75. Buenrostro J, Wu B, Chang H, Greenleaf W. ATAC-seq. A method for assaying chromatin accessibility genome-wide. Curr Protoc Mol Biol. 2015;109:21.29.1-9.

76. Ong CT, Corces VG. CTCF: an architectural protein bridging genome topology and function. Nat Rev Genet. 2014, 15(4):234-46.

77. Seiliez I, Froehlich JM, Marandel L, Gabillard JC, Biga PR: Evolutionary history and epigenetic regulation of the three paralogous pax7 genes in rainbow trout. Cell Tissue Res. 2015, 359(3):715-27.

78. Covelo-Soto L, Leunda PM, Perez-Figueroa A, Moran P: Genome-wide methylation study of diploid and triploid brown trout (*Salmo trutta* L.). Anim Genet. 2015, 46(3):280-88.

79. Baerwald MR, Meek MH, Stephens MR, Nagarajan RP, Goodbla AM, Tomalty KMH et al. Migration-related phenotypic divergence is associated with epigenetic modifications in rainbow trout. Mol Ecol. 2016, 25(8):1785-00.

80. Dann SG, Allison WT, Veldhoen K, Johnson T, Hawryshyn CW. Chromatin immunoprecipition assay on the rainbow trout opsin proximal promoters illustrates binding of NF-kappa B and c-jun to the SWS1 promoter in the retina. Exp Eye Res. 2004, 78(5):1015-24.

81. Vuori KA, Nordlund E, Kallio J, Salakoski T, Nikinmaa M. Tissue-specific expression of aryl hydrocarbon receptor and putative developmental regulatory modules in Baltic salmon yolk-sac fry. Aquat Toxicol 2008, 87(1):19-27.

82. Lo JH, Chen TT. CCAAT/enhancer binding protein beta2 is involved in growth hormone-regulated insulin-like growth factor-II gene expression in the liver of rainbow trout *(Oncorhynchus mykiss)*. Endocrinology 2010, 151(5):2128-39.

83. van Berkum NL, Lieberman-Aiden E, Williams L, Imakaev M, Gnirke A, Mirny LA et al.: Hi-C: A method to study the three-dimensional architecture of genomes. J Vis Exp. 2010;(39).

Macqueen et al. *Functional Analysis of All Salmonid Genomes (FAASG)*