

Functional Annotation of Salmonids Workshop Report

13 - 14th June, 2016 Toronto International Airport, Canada

EXECUTIVE SUMMARY

PARTICIPANTS AND OBJECTIVES

In 2009 the International Cooperation for the Sequencing of the Atlantic Salmon Genome (ICSASG)¹ embarked on a \$10M international project to sequence the genome of the Atlantic salmon. Following the publication of the Atlantic Salmon genome in April 2016, the ICSASG invited key international stakeholders to a workshop to discuss the opportunities for an international collaboration afforded by the availability of the reference salmon genome.

Participants from 19 institutions in ten countries were invited to attend based on their leadership in salmonid research and engagement in ICSASG sponsored International Conferences of Salmonid Biology (ICISB) meetings in 2012, 2014 or 2016. The following institutions were represented:

- University of Victoria (Canada)
- Genome British Columbia (Canada)
- Simon Fraser University (Canada)
- Instituto Tecnológico del Salmón (Chile)
- Universidad de Chile (Chile)
- Universidad de Santiago de Chile (Chile)
- University of Life Sciences (Norway)
- The Research Council of Norway (Norway)
- University of Tasmania (Australia)
- University of Alberta (representing FAANG)
- University of Turku (Finland)
- INRA (France)
- Matis (Iceland)
- University of Aberdeen (Scotland)
- University College Cork (Ireland)
- Roslin Institute (Scotland)
- University of Aberdeen (Scotland)
- USDA (USA)
- University of Washington (USA)

¹ The International Cooperation for the Sequencing of the Atlantic Salmon Genome (ICSASG) including: Genome BC, Research Council of Norway (RCN), Norwegian Seafood Research Fund (FHF), The Chilean Economic Development Agency – CORFO, InnovaChile, an Industry partners.

The objectives of for the workshop were identified as follows:

- Understand the state-of-the-art of salmonid genomics and associated tools.
- Understand the scope and operations of other functional annotation work being undertaken, especially that associated with terrestrial species (eg. the Functional Annotation of Animal Genomes initiative, www.FAANG.org).
- Discuss the need to improve traditional annotation, and to explore variation in transcribed and regulatory regions that could have a significant effect on the function of genes and salmonid biology.
- Enable networking and sharing of information on ongoing and planned research activities and interests that could align with work on functional annotation of salmonids.
- Initiate and develop scientific thinking regarding the opportunity to undertake a large-scale international collaboration on the Functional Annotation of All Salmonid Genomes (FAASG).
- Initiate engagement of funders of salmonid research in discussions around an international initiative on a FAASG activity.
- Collaboratively identify how the group can advance an international initiative on FAASG.

KEY MESSAGES

The following are key messages that arose during the workshop discussion. These themes are discussed in more detail in the body of the document and workshop presentations and background material in the Appendices.

- Salmonids are the focus of a great amount of research, however much of this data is housed in private databases and websites. This fragmentation along with lack of standards (e.g. transcriptomics and chromosome nomenclature) is limiting advances in salmonid research and development.
- There is interest in the salmonid community to take forward work on functional annotation of salmonids: to identify gene variants and their location and link this to their function or physical traits.
- The FAANG initiative provides a framework for possible collaboration on FAASG.
- Functional annotation work in salmonids should be harmonized or aligned with FAANG to the extent possible, however it is not clear whether FAANG core assays will work well for salmon, or whether new assays need to be developed.
- Given the breadth of salmonid species (70-200) collaboration on the functional annotation of salmonids could provide a depth and richness to understanding of salmonid biology not possible for commercial terrestrial species.
- There is a need to engage the broader salmonid community in thinking about functional annotation of salmonids

- Scientists agreed on the need for an international initiative to support functional annotation of salmonids and agreed on a plan of work to lay the foundation for the initiative and engage the larger salmonid community in its development and operation.
- Industry engagement is needed and initiatives will be taken.

NEXT STEPS

The group agreed to activities to (i) develop a Functional Annotation of All Salmonid Genomes (FAASG) Initiative and (ii) engage the international salmonid community and industries. Specific next steps for each of these are below.

- Development of a FAASG Initiative
 - Establishment of FAASG identity and connection to FAANG
 - Establishment of framework for the FAASG initiative
 - Development and review of a pilot project validating the FAANG Core Assays
 - Exploration of the EU-FAANG COST action for a possible broader engagement in the EU-FAANG COST action
- Engagement of international salmonid community in FAASG
 - Development of a website
 - Development of a white paper
 - Outreach to the larger salmonid and FAANG communities
 - An expanded FAASG Workshop at the Plant and Animal Genomics Meeting, January 2017

TABLE OF CONTENTS

EXECUTIVE SUMMARY	1
Participants and Objectives	1
Key Messages	2
Next Steps	3
WORKSHOP SUMMARY	4
Background.....	4
Participant Selection	5
Workshop Objectives.....	5
Inaugural workshop	5
Agenda	5
I. SETTING THE SCENE: THE STATE OF SALMONID RESEARCH.....	6
Overview of ICSASG and background to the meeting (Ben Koop)	6
II. SETTING THE SCENE: FUNCTIONAL ANNOTATION.....	7
FUNCTional Annotation of Animal Genomes (Graham Plastow)	7
III. TOWARDS FUNCTIONAL ANNOTATION OF SALMONIDS	8
The salmonid genomes as a basis for coordinated Functional annotation activity (Sigbjorn Lien)	8
Scientists' roundtable (Willie Davidson).....	9
IV. OPEN DISCUSSION.....	9
towards a functional annotation of salmonids initiative (Rachael Ritchie & Steinar Bergseth).....	9
V. DISCUSSION ON NEXT STEPS	12
APPENDIX A: TORONTO WORKSHOP Agenda.....	13
APPENDIX B: Presentation- Ben Koop	14
APPENDIX C: Presentation – Graham Plastow	15
APPENDIX D: Presentation – Sigbjorn Lien	16
APPENDIX E: PARTICIPANTS	17

WORKSHOP SUMMARY

BACKGROUND

In 2009, Norway, Chile and British Columbia Canada agreed to collaborate in the sequencing of the Atlantic salmon genome as to provide a reference genome to advance our understanding of this culturally, environmentally and economically important species. The result was the formation of the International Cooperation to Sequence the Atlantic Salmon Genome (ICSASG), a \$10M USD investment by the partners that resulted in a landmark paper published in Nature in April 2016. Along the way, the project engaged researchers from the US, and spawned three International Conference or Integrative Salmonid Biology (ICISB) in 2012, 2014 and 2016 in Oslo, Vancouver and Puerto Varas respectively.

At the ICISB2016 meeting, the ICSASG funders and scientists invited the international community to come together and consider new work based on salmonid knowledge, reference sequences and stakeholder's needs, that could be constructively taken forward in an international collaboration. Discussion identified functional annotation as a possible focus for new work and the ICSASG leaders agreed to convene a workshop in Toronto with key international stakeholders discuss the idea.

PARTICIPANT SELECTION

The ICSASG were keen to involve participants from as many new countries as possible including from Australia, Finland, France, Iceland, Ireland, Northern Ireland, Scotland and the US. Scientists and potential funders from these countries were invited.

The number of participants was kept intentionally small to allow time and scope for productive discussion (Appendix E).

Scientists were identified based on expressed interest in functional annotation, track record in collaborative research, and leadership in the community.

Participation from the three ICSASG countries was prioritized to scientists knowing the history of the sequencing project and the initial FAASG thinking. A broader involvement will be done.

WORKSHOP OBJECTIVES

The workshop aimed to initiate and develop scientific thinking regarding the opportunity to undertake a large-scale international collaboration on the functional annotation of salmonids. Such work is now possible given existence of a high quality reference genome and might be supported by scientists, funders, industry and others interested in increasing our understanding of salmonids and using that information to improve aquaculture, conservation and fisheries management for over 70 economically, culturally and environmentally important salmonid species worldwide.

The workshop provided a forum for presentation and discussion of the current trends and technologies in functional annotation, drawing on experiences from the Functional Annotation of Animal Genomes (FAANG) initiative. These developments provided a framework for discussions regarding the type of scientific research and developments that might be undertaken on an international level.

INAUGURAL WORKSHOP

AGENDA

The inaugural workshop was focused on setting the scene for discussion of an international effort focused on functional annotation of salmonids. The agenda (Appendix A) had four parts:

- **Part I: Setting the scene: the state of salmonid research** Ben Koop provided the history of the ICSASG and the scientific outcomes of that work leading to the starting point for new work to be discussed at the workshop.
- **Part II: Setting the scene: Functional Annotation** Graham Plastow provided an overview of the Functional Annotation of Animal Genomes (FAANG) initiative.
- **Part III: Towards functional annotation of salmonids** Sigbjørn Lien provided some details on available salmonid genomic resources and a justification for functional annotation of salmonid species. Willie Davidson then chaired a roundtable of participant research interest and alignment related to functional annotation of salmonids.
- **Part IV: Open discussion and agreement on next steps** Participants had an opportunity for open discussion around the opportunity for collaborative work on the functional annotation of salmonids and how that could be taken forward. Mid way through discussion participants divided into groups for deeper discussion on some key next steps that should be considered in developing an initiative for functional annotation of salmonids.

I. SETTING THE SCENE: THE STATE OF SALMONID RESEARCH

OVERVIEW OF ICSASG AND BACKGROUND TO THE MEETING (BEN KOOP)

Ben provided an overview of the ICSASG collaboration as an example of how international collaboration can be used to tackle large and complex problems (Appendix B). The ICSASG collaboration was quite unique in that funders had a strong role in developing the project and engaging the best scientists and tools regardless of their location. The project was successful in reaching its goal of generating a high quality reference genome and, with this resource in-hand, we are now in the enviable position of considering new work that could be advanced in collaboration with the broad scientific community.

There are approximately 70-200 salmonids species, and more than 100,000 papers on salmonids on the Web of science with 50% focused on fisheries science. In 2015 alone, close to 7000 papers were published in the field, highlighting the high level of research activity and interest in salmonids. This number of papers dwarfs most terrestrial production species and model species such as zebra fish. To date three full genomes exist, with more in the pipeline given the available resources. Despite this activity, public database is lacking--containing less than a 'page' for teleosts--and it is presumed that much data exist in private databases/ websites.

Review of the field revealed lack of standardization and fragmentation of efforts that might hinder advancement of the field. For instance, syntenic chromosomes have no consistent nomenclature across different species. Different standards are also being used when studying transcribed regions of the genome (transcriptome). This is especially important given the duplicated nature of salmonids. There is a pressing need to standardize use of existing tools and harmonization of data so that key questions about salmonid biology can be answered efficiently based on shared data sources, including relating to immune function / health, olfaction, smoltification etc. These questions underlie effective management of wild populations and sustainable production of

farmed salmonids. To answer these questions we need to link genes variant (and location) with physical traits to understand salmonid biology.

II. SETTING THE SCENE: FUNCTIONAL ANNOTATION

FUNCTIONAL ANNOTATION OF ANIMAL GENOMES (GRAHAM PLASTOW)

Graham provided a history and overview of FAANG. FAANG is a coordinated, international action focusing on terrestrial species, that began in 2014, as an 'offshoot' of the human ENCODE project (Appendix C). Its goal is to accelerate 'genome to phenome', that is, the ability to understand animal phenotype from genomic information. The initiative is still very much a grass roots organization but its profile and impact in the animal genome world is growing.

FAANG encourages the international community to come together to collaborate on functional annotation of, and across, animal species. To do this, FAANG promotes use of standard core assays, coordination and facilitation of data sharing, and the development of necessary biobanking and analysis infrastructure. Individuals are welcome to contribute in FAANG providing they adhere to the Guiding principles of open access and immediate release of data to the community (under the 'Toronto Principles' and the 'Fort Lauderdale Principles').

According to the FAANG white paper (2015, http://www.faang.org/white_paper distributed to participants prior to the meeting) functional annotation requires a high quality reference genome, support of the research community, a critical mass of investigators, as demonstrated by expression of interest and willingness to use core assays, and a common data-sharing infrastructure. Accordingly most work to date has focused on domesticated animal species.

For each species, the first phase of the FAANG project focuses on sampling biological replicates representing a limited number of specific biological states to maximize comparisons across species. Where possible, animals with minimal genetic diversity within a species will be sampled in a limited range of developmental phases (eg neonatal and mature phases). Core tissues are archived and core assays are established for the species in this first phase.

In the second phase, additional samples (with greater genetic diversity) are collected and analysed to answer more discrete biological questions, with additional assays being developed as needed to address specific needs and research interests. Today, FAANG projects on beef and dairy cattle, pigs, sheep, and chickens exist with researchers working together to coordinate funding for discrete projects. FAANG researchers are also pursuing additional sources of funding through European and US agencies. FAANG participants meet to exchange information via steering committee calls and participation in four thematic working groups and an annual general meeting, often held on the margins of another meeting (e.g. PAG).

Projects are being built in 'bricks' with different partners pursuing different funding mechanisms. In Europe, a COST action is being pursued, in Australia, Ben Hayes is seeking funding for core assays on cows, and in the US, the USDA-AFRI has released an RFP to support functional

annotation in terrestrial species. At present the FAANG community must balance work with funded projects on functional annotation with adherence to the FAANG initiative and contributing to the development of FAANG as part of daily routines. The FAANG White paper and Graham's presentation fed into subsequent workshop discussions.

- **Q&A/Comments**

In an active dialogue many questions and comments were directed towards understanding the FAANG initiative and how the salmonid initiative can benefit from the experiences made. Discussion then focused on the need for a plan to understand how a salmonid initiative might develop e.g.: including a salmonid FAASG / FAANG structure, phasing of its setup and development, standardized recordings of phenotypes often owned by industry, expression profiles, involvement of physiologists, how to share all types of data to avoid duplications and with the right QC, gene annotation to be completed, etc...

III. TOWARDS FUNCTIONAL ANNOTATION OF SALMONIDS

THE SALMONID GENOMES AS A BASIS FOR COORDINATED FUNCTIONAL ANNOTATION ACTIVITY (SIGBJORN LIEN)

Sigbjørn presented the need to improve traditional annotation, to explore variation in transcribed and regulatory regions that could have a significant effect on the function of genes and salmonids. Functional annotation, is also important for understanding how genome function evolves after whole genome duplication. The topic is best tackled by the salmonid community, especially given the complexity arising from sub-functionalization of portions of the duplicated genome. Lien made the case for generation of salmonid 'Atlases' (SalATLASes) that contain functional information on important traits.

SalATLASes could be generated through common garden experiments with multiple salmon species, including pike as a control, to define standards. He proposed a range of tools to undertake comprehensive, multi-omic surveys of genome regulation (including RNA-Seq, Histone Mod Marks, ATAC-seq). Lien suggested adopting best practices from FAANG to standardize protocols and generate common data infrastructure, and he advocated for releasing experimental and metadata rapidly in an open access manner. Lien described pilots that might be undertaken including:

- NutATLAS:- metabolic functions in domesticated and wild salmonids
- DevATLAS:-developmental change - early developmental stages
- ImmuneATLAS – Immunity change
- SeasonATLAS – Seasonal change

Lien suggested the pilots, and larger projects, could bring together specialists from the salmonid community to focus on each area, with core assays being undertaken in core facilities.

- **Q&A/Comments**

Project's setup need to be discussed as there are certain challenges experimenting with live fish. Experiences are on hand and well thought designs can be made. StressATLAS was also proposed.

SCIENTISTS' ROUNDTABLE (WILLIE DAVIDSON)

Willie moderated a scientists' roundtable where participants provided a brief statement of ongoing research and research interests that might align functional annotation of salmonids work:

Summary:

- Scientists participating in the meeting are working on a range of salmonids including Atlantic salmon, Arctic Charr, Brown trout, Rainbow trout, Greyling and also pathogens like salmon lice and others. Research interests ranged from salmonid biology to traits involved in production and conservation including life history adaptation, age at sexual maturation, homing instincts, disease resistance, thermal tolerance, nutrition etc. Participants are using a range of genomic and genetic tools for this work including genetic mapping, systems biology, SNP genotyping, sequencing, linkage mapping, rad-seq, genotyping by sequencing, RNA seq, quantitative genetics, mtDNA analysis and cell line models. Few of the participants were using FAANG core assays other than RNA-seq.

IV. OPEN DISCUSSION

TOWARDS A FUNCTIONAL ANNOTATION OF SALMONIDS INITIATIVE (RACHAEL RITCHIE & STEINAR BERGSETH)

Sensing participant interest in working together on functional annotation of salmonids, Rachael and Steinar moderated a discussion on how this might be taken forward. They asked if the community can begin immediately by joining FAANG, agreeing to principles and working towards standardization? Whether salmonid researchers could join FAANG working groups and working toward these standards? Participants were also asked to comment on gaps that might need to be addressed including: the need for additional core assays, where and how samples might be kept (biorepository), whether an appropriate platform for putting /accessing data exists, the need for standardization and how an initiative might operate/be governed, and whether international collaboration on common garden experiments would be useful, or how international collaboration on strategic themes (to reduce duplication) (eg ATLAS idea) might work.

- **Q&A/Comments**

Technical and assay-specific comments and questions concerned antibody availability for methylation in fish DNA, which standard assays will be most suitable for salmonids, the need for an RNA-seq baseline, what are the key common phenotype(s) to work on for all salmonids, the need for double haploids as a tool, understand if/what constraints there might be deposition of data types, etc... Several comments pointed to the usefulness of connecting to FAANG to take advantage of their experiences and possible core assays, but also give something back. Salmonids have their specificities and unique possibilities so a strong FAASG group should be built. What are the overarching questions to handle within a FAASG initiative?

Following a wide-ranging discussion three areas requiring more detailed discussion were identified:

(A) **Engagement with FAANG:** how closely should the salmon work coexist with FAANG, recognizing the best practices developed by FAANG, and ENCODE before it, but balancing some of the unique features of salmonids and the connectedness of the community?

(B) **Engagement with the broader salmonid community:** Can preparation of a white paper be used to lay out the opportunity for functional annotation in salmonids and engage the broader salmonid community? What else can be done to engage the community?

(C) **Development of a pilot project:** what work needs to be done now to help frame future, coordinated functional annotation work for salmonids?

Participants were invited to separate into groups to discuss each area in more detail before presenting their thinking to the group in plenary for final discussion and identification of next steps:

(A) ENGAGEMENT WITH FAANG

The group proposed an approach to engaging with FAANG while also developing some FAASG specific structures. Current thinking is included below although it was agreed in plenary that final decisions are best made once more information on FAANG and the core assays is understood and the international community is engaged further.

Integration with FAANG

- Ready consensus on joining FAANG
- Some discussion on “ownership” of data under Toronto agreement, clarified by Graham. Researchers should contact originators and clarify plans for publication
- There are two immediate opportunities for outreach with FAANG
 - Presentation of the FAASG initiative at the FAANG Symposium (July 23) held in conjunction with the International Society for Animal Genetics meeting in Utah (in Plenary, Caird offered to attend)
 - In Chile, large meeting with funders in near future, Patricia and Rodrigo will attend

FAASG Working Groups to connect to FAANG working groups were proposed:

1. Integration with FAANG (funders) working group
2. Integration of existing data sets working group
3. Pilot study working group
4. Life history trait working group
(includes anadromy, maturation, return timing, spawn timing)
5. Production phenotypes of animal production working group
(includes nutrition, reproduction, immune systems, stress)
6. Genomic tools and assays

● *Comments in plenary included:*

WG 2:

DRAFT

- *Data sets already exist*
- *Early effort should be focused collating transcriptomes across species*
- *We should examine whether data can be readily linked to FAANG*
- *Should we look to the PhyloFish model? Or develop our own equivalent to this model, that speaks to FAANG and PhyloFish*
- *Can we mine the existing RNAseq data and perform functional annotation of the SNPs?*

WG 4 and 5:

- *FAANG encourage “champions” for traits within each group*

WG 6:

- *Group could examine new tools, and screen existing FAANG tools*

We need some scientists to link to FAANG working groups/steering group

(B) ENGAGEMENT WITH THE BROADER SALMONID COMMUNITY

The group discussed the importance of a white paper as a high profile way to engage the international community. The paper will in essence represent the will and interests of the scientists at the Toronto meeting and would be co-written.

In subsequent plenary discussion it was discussed that the paper might contain the following information:

1. Salmonids and their importance for food security and ecology and evolution
2. Similarities between the phenotypic traits of interest for aquaculture and ecology and evolution
3. Progress in the field since the 2010 white paper (success stories, what are the remaining questions?)
4. How can functional annotation help to address these questions? What are the potential benefits?
5. Rational for functional annotation and linking with FAANG (including lessons from FAANG, what salmonids can give back to FAANG, model of for other fish species and other species of ecological and evolutionary interest)
6. Differences for salmonids compared to current FAANG

The paper would also aim to integrate the following ideas / concepts:

- Key FAANG differences for salmonids
- Recent domestication, recent adaptation
 - large numbers of offspring allow more efficient genomic selection
- Numbers of offspring, family sizes. External fertilization, ability to create the families/crosses needed. Use of double haploids.
- Standardised phenotypes via scale readings
- Well annotated genomes for closely-related species
- Relevance for aquaculture and industry, but also for conservation and management, ecology and evolution.
- Genome duplication
- Recombination patterns- importance for evolution and production/selection
- Key questions and challenges

- increased emphasis on phenotypes (ecological and production)- morphological and life history, also response to treatment

In plenary, the group discussed the timing of the paper and it was agreed that it could be developed in parallel with related activities to promote the idea and engage the community. Contributors for different sections of the white paper were also discussed.

(C) DEVELOPMENT OF A PILOT PROJECT

This group addressed the question of whether established FAANG core assays would work for hyper-methylated salmonid species. The group discussed an approach to testing this. Three priority species were identified: Rainbow trout, Atlantic salmon and Pike.

The group discussed different treatments, and the use of cell lines vs fish tissues. The group agreed that the work should be done by labs, possibly FAANG labs, already proficient in the assays, rather than developing them in-house. They agreed that the pilot should be conducted as soon as possible and suggested it could be done for between \$7500-30000 USD. The ICSASG consortium indicated possibility of funding a pilot pending review of an appropriate proposal.

- ***Comments in plenary included:***

How white paper and pilot experiments should be timed? Probably best to do in parallel so info on core assays can be part of the white paper. Cell lines vs tissue? Not many cell lines exist for salmonids or pike. Timeline for pilot and whitepaper?

V. DISCUSSION ON NEXT STEPS

A number of recommendations were made by the breakout groups and determined, by the larger group in plenary session, to support future work on functional annotation of salmonids. The activities can be grouped into two broad categories:

(i) work to determine to what extent FAANG and the core assays are applicable to salmonids (including participation in FAANG and development and execution of a pilot proposal)

(ii) engagement of the broader salmonid community in the functional annotation of salmonids work (including development of a white paper, promotion at a scientific meetings, emails to scientists and a website).

A workplan will be developed over the coming months with feedback from the salmonid community.

<http://www.faasg.org/wp-content/uploads/2016/08/Appendix-A.pdf>

<http://www.faasg.org/wp-content/uploads/2016/08/Appendix-B.pdf>

<http://www.faasg.org/wp-content/uploads/2016/08/Appendix-C.pdf>

<http://www.faasg.org/wp-content/uploads/2016/08/Appendix-D.pdf>

The meeting was arranged at relatively short notice so not all invitees were able to join. The list of attendees is as follows:

- Ben F Koop, University of Victoria (Canada)
- Rachael Ritchie, Genome British Columbia (Canada)
- Willie Davidson, Simon Fraser University (Canada)
- Alfredo Tello, Instituto Tecnológico del Salmón (Chile)
- Patricia Iturra, Universidad de Chile (Chile)
- Rodrigo Vidal, Universidad de Santiago de Chile (Chile)
- Jon Olav Vik, University of Life Sciences (Norway)
- Sigbjørn Lien, University of Life Sciences (Norway)
- Steinar Bergseth, The Research Council of Norway (Norway)
- Barbara Nowak, University of Tasmania (Australia)
- Graham Plastow, University of Alberta (representing FAANG)
- Craig Primmer, University of Turku (Finland)
- Yann Guiguen, INRA (France)
- Kristinn Ólafsson, Matis (Iceland)
- Philip McGinnity, University College Cork (Ireland)
- Daniel MacQueen, University of Aberdeen (Scotland)
- Ross Huston, Roslin Institute (Scotland)
- Sam Martin, University of Aberdeen (Scotland)
- Caird Rexroad, USDA (USA)
- Kerry Naish, University of Washington (USA)