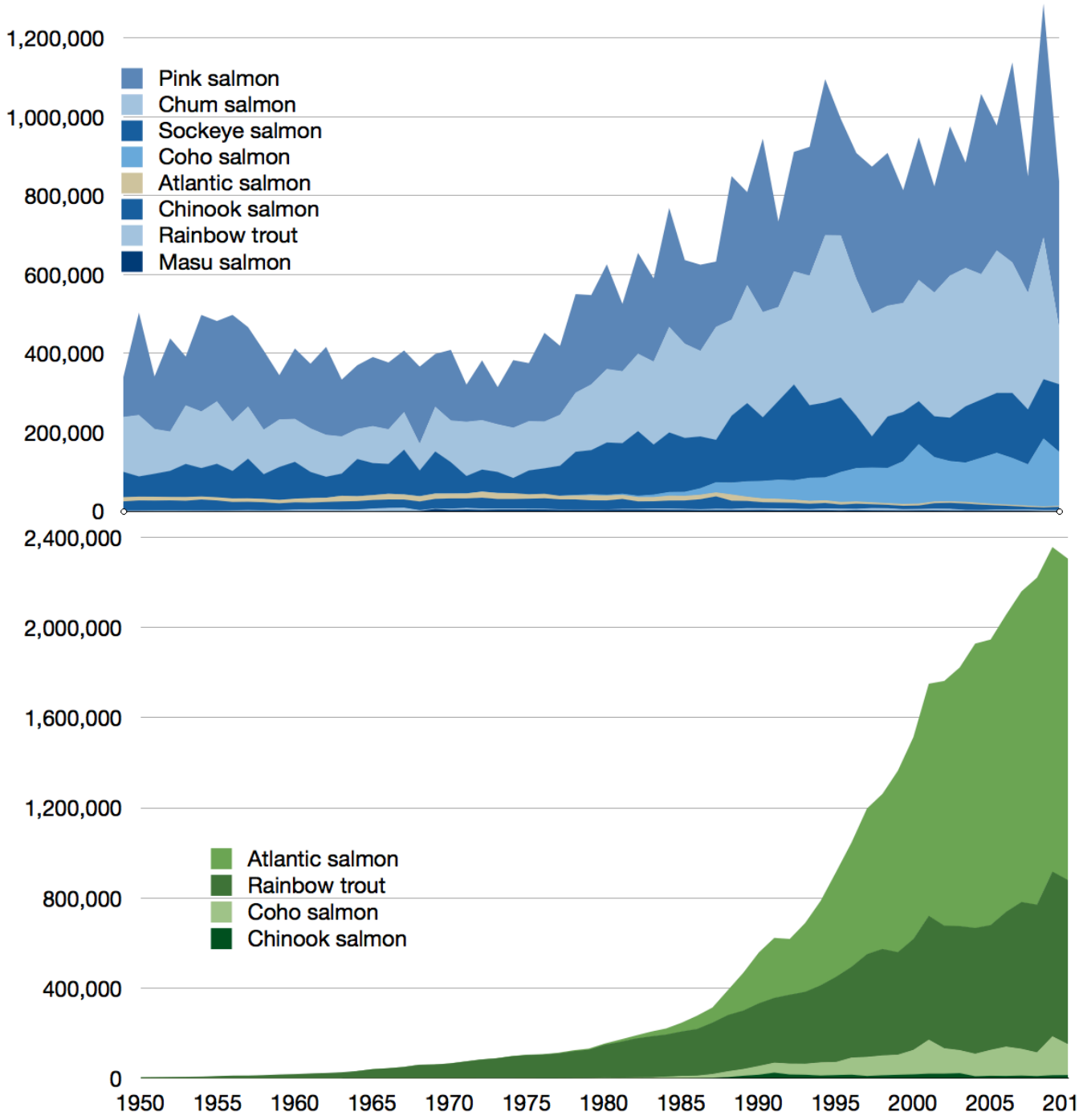
Functional Annotation of All Salmonid Genomes (FAASG)

Approach: Coordinate the international salmonid community to acquire, standardize and share data for comprehensive mapping and characterization of the functional elements of salmonid genomes.

Salmonid Fisheries and Aquaculture



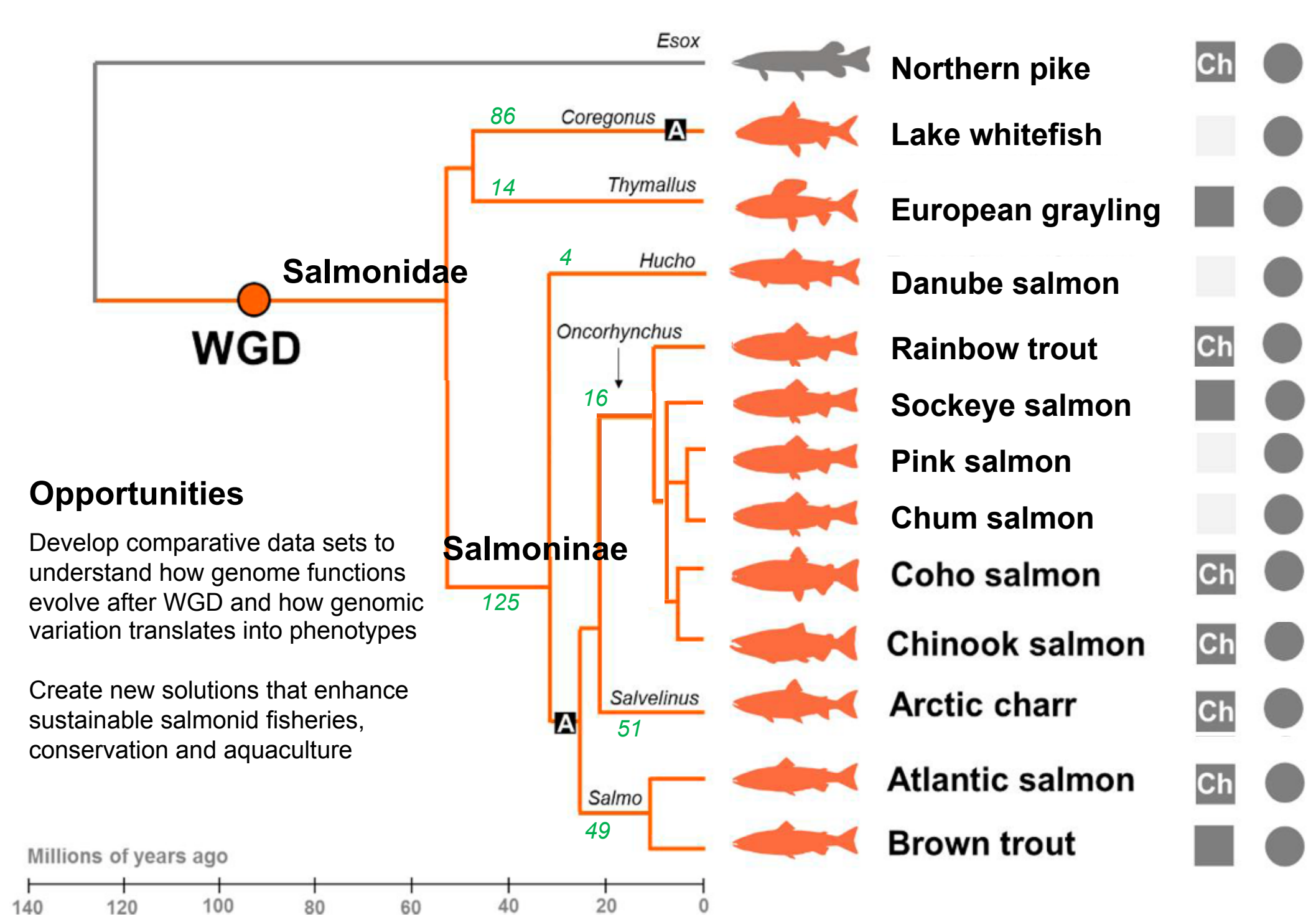
Wild fisheries – commercial capture of all wild salmon species FAO

Aquaculture production of all salmon species FAO

2016 – $15B AS

Scientific Publications by Fish Species and Subject

Subject Area ***Publications*** *(1900-2016)*	***Salmonids***	***Catfish***	***Halibut & Flounder***	***Atl. Cod & Haddock / Hake***	***Tuna & Mackerel***	***Carp***	***Tilapia & Cichlids***	***Zebrafish***	***Stickleback***	***Medaka***
Total	**101k**	**17.6**	**11.9**	**18.2**	**11.3**	**32.2**	**16.3**	**35.0**	**4.8**	**6.3**
Fisheries	**34.1**	**5.8**	**4.7**	**6.5**	**3.4**	**6.9**	**4.2**	**1.1**	**0.6**	**0.5**
Marine/FW Biology	**24**	**3.3**	**3.9**	**6.3**	**2.5**	**4.8**	**3.0**	**1.6**	**0.9**	**0.8**
Zoology	**9.7**	**2.7**	**1.0**	**1.0**	**0.6**	**3.3**	**3.0**	**2.3**	**1.4**	**0.9**
Biochemistry	**9.7**	**1.5**	**1.1**	**1.2**	**0.6**	**2.6**	**1.5**	**5.6**	**0.3**	**0.9**
Environment	**9.9**	**1.2**	**1.0**	**0.8**	**0.6**	**2.3**	**1.0**	**1.7**	**0.2**	**1.0**
Toxicology	**7.4**	**1.0**	**0.8**	**0.4**	**0.2**	**1.5**	**0.7**	**2.5**	**0.1**	**1.2**
Ecology	**6.4**	**0.7**	**0.6**	**1.8**	**0.7**	**1.5**	**1.4**	**0.2**	**1.4**	**0.1**
Food Science Technol.	**3.8**	**0.8**	**0.3**	**1.8**	**2.1**	**0.9**	**0.6**	**0**	**0**	**0**
Immunology	**3.4**	**0.9**	**0.6**	**0.3**	**0.1**	**1.4**	**0.3**	**0.9**	**0**	**0.1**
Genetics	**2.9**	**0.5**	**0.3**	**0.3**	**0.1**	**1.2**	**0.9**	**3.3**	**0.6**	**0.7**
Cell Biology	**2.0**	**0.3**	**0.3**	**0.2**	**0.1**	**0.6**	**0.3**	**4.5**	**0.1**	**0.6**
Agriculture	**1.2**	**0.3**	**0.1**	**0.0**	**0.3**	**0.5**	**0.4**	**0.1**	**0**	**0**
Development	**0.8**	**0.1**	**0.0**	**0.0**	**0.0**	**0.2**	**0.2**	**7.1**	**0.1**	**0.7**
Conservation	**1.1**	**0.3**	**0.1**	**0.2**	**0.1**	**0.3**	**0.1**	**0**	**0.1**	**0**
Est Value US$, 2006 (fisheries +aquaculture)*	**12.3B**	**0.4B**	**2.8B**	**6.1B**	**10.5B**	**19.5B**	**3.5B**			
Est MT*	**3**	**0.6**	**1.0**	**6.1**	**6.5**	**21.2**	**3.1**			



Opportunities

Develop comparative data sets to understand how genome functions evolve after WGD and how genomic variation translates into phenotypes

Create new solutions that enhance sustainable salmonid fisheries, conservation and aquaculture

 # species: FishBase 225 total

Macqueen *et al. BMC Genomics* (2017) 18:484
DOI 10.1186/s12864-017-3862-8

BMC Genomics

EDITORIAL Open Access

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

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], The FAASG Consortium

FAASG Vision:

To provide solutions for salmonid conservation, sustainable fisheries, and aquaculture through an improved understanding of salmonid biology.

FAASG Mission:

To coordinate the salmonid community to standardize and share data for comprehensive mapping of the functional elements of salmonid genomes.

Core Principles

To understand the functional elements of all salmonid genes and genomes.

Modeled on principles agreed on by similar initiatives including FAANG and ENCODE including:

- **Collaboration to define experimental, meta data, and bioinformatics standards**
- **Ensuring experiments conducted adhere to agreed standards**
- **Timely and open access release of data**

	FAANG	FAASG
Initiated	2013	2016
Focus	Terrestrial animals (commercial)	Salmonids (industry and conservation)
Lead Countries	USA / Scotland / France	Canada / Norway / USA / Scotland
White Paper	2015	2017
Governance/ Working Groups	Yes	Yes
Core Assays	Yes	in progress
Metadata	Yes	in progress
Bioinformatics	Yes	in progress
Phenotypes	Yes	in progress
Funding	Core funding for Infrastructure (EU) and (ad hoc) project funding	ICSASG, National funds, and possibly EU funds
Origin	Community driven	ICSASG (funder) driven

ASA

Metadata

BioInf

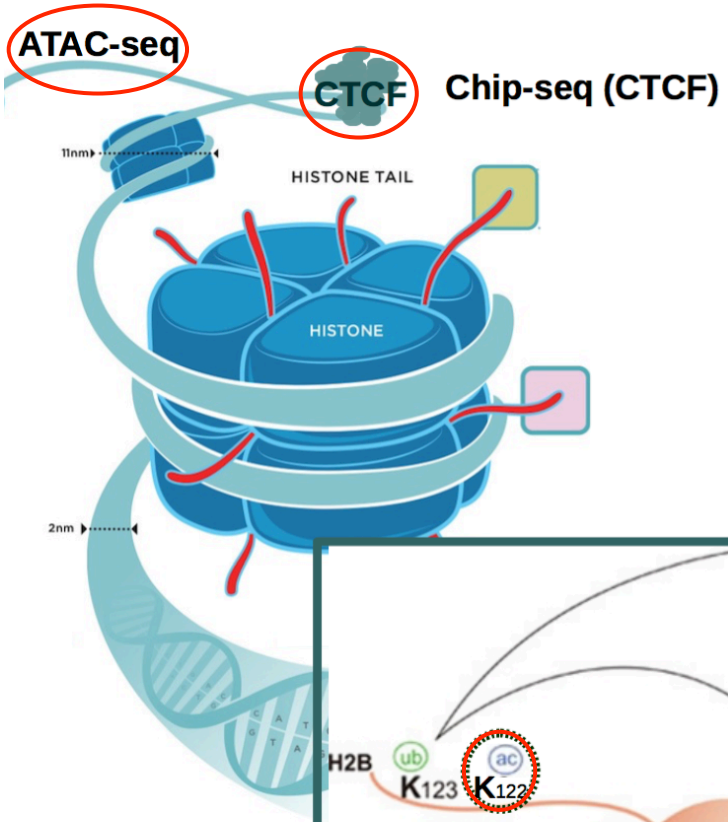
Phe

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

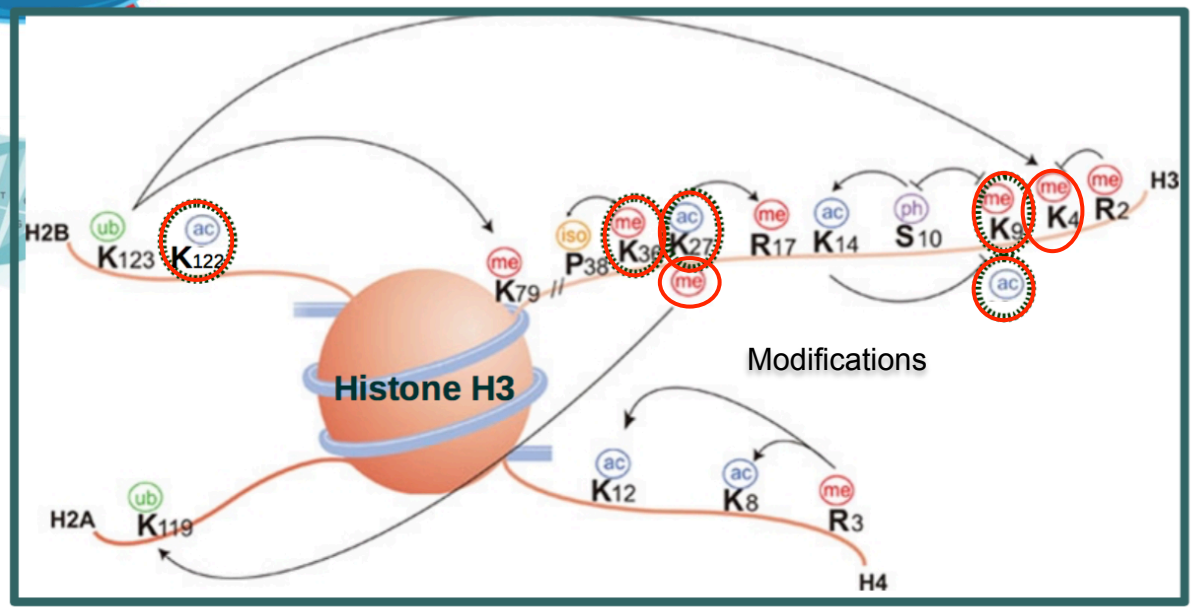
Genome Annotation

Epigenetic Assays (3 coho)



H2BK122 ac
H3K27 ac
H3K9 ac
H3K36 me
H3K9 me
H3K27 me
H3K4 me

ATAC-seq ➔ openness
CTCF – Chip-seq ➔ promoter
DNA- methylation (hyper/hypo)
RNA-seq ➔ expression



Liver Tissue – Test system
➔ 50% hepatic cells

- **triglycerides & cholesterol**
- **Glycogen synthesis**
- **Clotting factors**
- **Detoxification**

Regulation of chromatin by histone modifications Andrew J Bannister & Tony Kouzarides
Cell Research (2011) **21**, 381–395 & Liwang_Cui2/publication/44583139/figure/fig1/AS:2775704

FAASG models - what are the priorities and challenges?

Tissue 'Atlas' – underpinning baseline to many phenotypes. Challenges include heterogeneity in cell-type, standardizing effects of ontogeny across species, choice of populations/strains …

Developmental 'Atlas' – underpinning to later phenotypes. Challenges include standardizing across species (easiest during embryogenesis), and loss of tissue-specific signal …

Immune 'Atlas' – potential to do standardized challenges with immune mimics. Issues and ideas?

Clonal and selected strains – reduced effect of genetic variation and shifted phenotypes (e.g. disease resistance)

Life-stage transitions – sexual maturation and smoltification. Many challenges in terms of standardization across studies.

Cell lines and primary cell cultures – benefits for core assays. Functional manipulations possible

Supporters

The Research Council of Norway

FHF
FISKERI- OG HAVBRUKSNÆRINGENS FORSKNINGSFOND

USDA

InnovaChile
CORFO

CONICYT
COMISIÓN NACIONAL DE INVESTIGACIÓN CIENTÍFICA Y TECNOLÓGICA
GOBIERNO DE CHILE

BBSRC
bioscience for the future

Genome BritishColumbia
Leading ▸ Investing ▸ Connecting

FAASG.org

Historical Overview

ICSASG

- MOU signed to initiate the "**International Collaboration to Sequence the Atlantic Salmon Genome (ICSASG)**" – April 2009
- International Cooperation Agreement (ICA) signed and $10M committed – October 2009
- Genome Biology publication – September 2010
- Reference genome publication in Nature – April 2016

Sequence

- Constructed using Sanger reads (12.2M) from plasmids, phosmids and BACS, as well as Illumina reads (800M)
- Consists of 555,000 contigs representing 2.43Gb sequence
  - Phase 1 – Sanger-based sequencing completed February 2011. First assembly released to GenBank – December 2011
  - Phase 2 – Illumina & PacBio sequencing completed December 2013. Reference assembly released to GenBank – March 2014

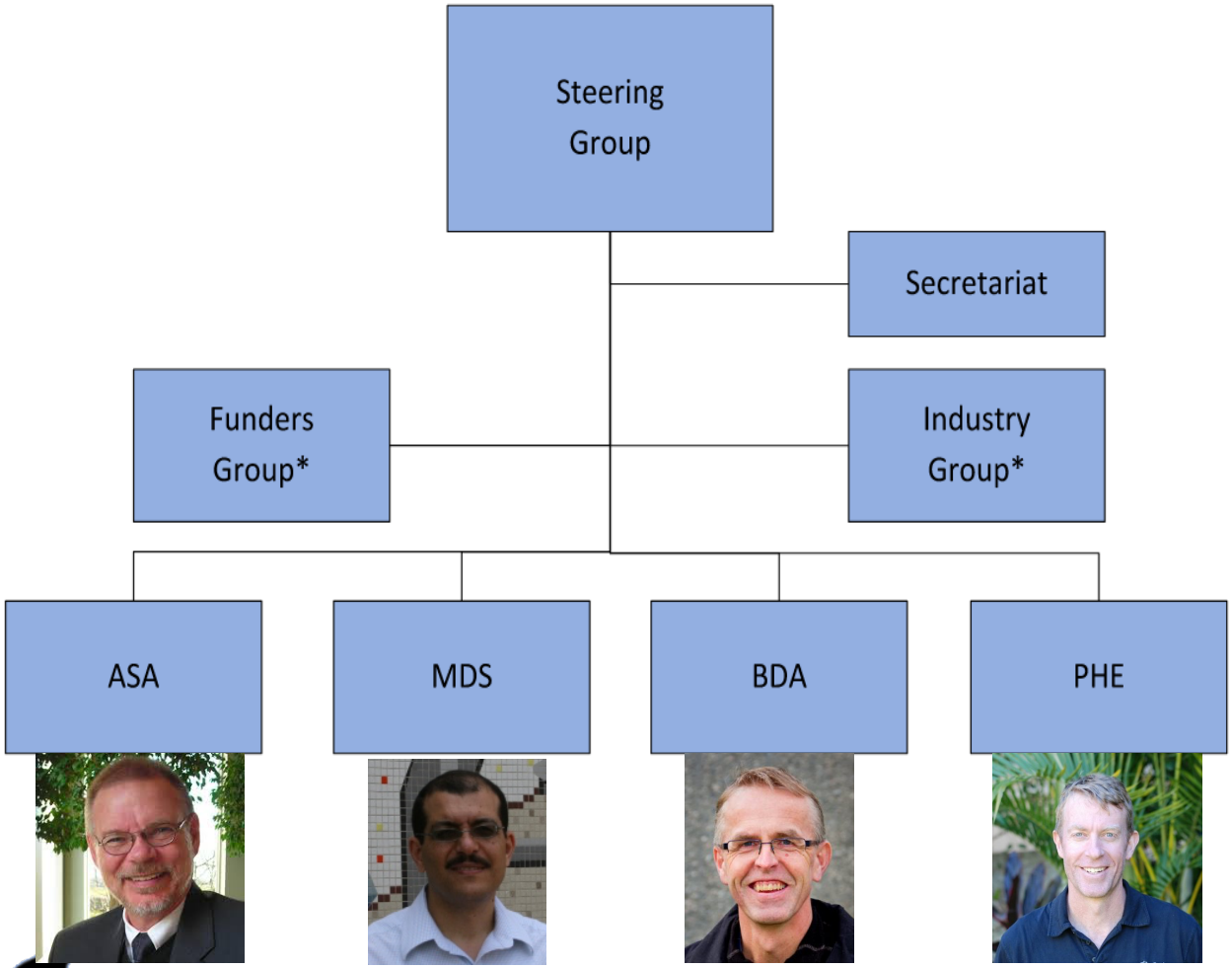
ICISB

- First International Conference on Integrative Salmon Biology (ICISB 1) – Oslo, Norway – June 2012
- Second International Conference on Integrative Salmon Biology (ICISB 2) – Vancouver, Canada – June 2014
- Third International Conference on Integrative Salmon Biology (ICISB 3) – Puerto Varas, Chile – April 2016

FAASG

- "**Functional Annotation of All Salmonid Genomes**" Initiative
  - Idea originated during initial discussions at the ICISB 3, Chile and initiated at the subsequent "Toronto Workshop" – June 2016
  - Workgroups established and the Vision and Mission documented, and a website was created to disseminate information, and allow for collaboration at the "San Diego Workshop" – January 2017
  - BMC Genomics publication – June 2017
- **Current and Next Steps**: (i) core-assay development infrastructure establishment, and (ii) validation of FAASG platform

Governance and Operations Structure



Working Groups

- Animals, Samples and Assays (**ASA**)
- Bioinformatics and Data Analysis (**BDA**)
- Metadata and Data Sharing (**MDS**)
- Phenotyping (**PHE**)

- Steering Group
- Secretariat

- Funders – still to be established
- Industry – still to be established

FAANG core assays	Assay target	Example related assays used in salmonids; assay target
Transcribed loci		
RNA-sequencing, stranded protocols [52, 68]	Transcriptome, strand polarity retained	RNA-seq, stranded protocol; transcriptome [69] RNA-seq; double stranded; transcriptome [65-67] RNA-seq; double stranded, miRNA [70] RNA-seq; double stranded; Long non-coding RNA [71] RNA-seq; double stranded; Large intergenic non-coding RNAs [72]
Chromatin accessibility and architecture		
Assay for transposase-accessible chromatin sequencing (ATAC-seq) [[57, 73]	Regions of open chromatin, localization of nucleosomes in regulatory sites and positions of DNA-binding proteins	No published examples
DNaseI footprinting [58]	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 [59]	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 H3 lysine 4 monomethylation (H3K4me1), marks enhancers and other distal elements, and is enriched downstream of transcription start sites	Chromosome immune precipitation; relationship between modified histones and gene expression [75]
Additional FAANG Assays	**Assay target**	**Related assay used in salmonids; assay target**
DNA methylation, genome-wide analysis of 5-methylcytosines, nucleotide level resolution [55]	Epigenetic mark and regulator of gene expression	Methyl-sensitive AFLP, global methylation changes [76-79] Bisulphite sequencing, nucleotide-level resolution [80]
ChiP-seq assays [62] for sequences bound by specific proteins	Transcription factor binding sites	Chromosome Immune Precipitation Assay, regulation of transcription [81-83]
Genome conformation; Hi-C [61, 84] for chromosomal conformation capture	Identify distal chromatin elements that are brought together through 3D chromosomal folding	No published examples