

OPEN LETTER

Sequencing the genome of the Atlantic salmon (*Salmo salar*)

William S Davidson^{1*}, Ben F Koop², Steven JM Jones³, Patricia Iturra⁴, Rodrigo Vidal⁵, Alejandro Maass⁶, Inge Jonassen⁷, Sigbjorn Lien⁸ and Stig W Omholt⁸

Abstract

The International Collaboration to Sequence the Atlantic Salmon Genome (ICSASG) will produce a genome sequence that identifies and physically maps all genes in the Atlantic salmon genome and acts as a reference sequence for other salmonids.

Economic, societal and scientific importance of salmonids

The family Salmonidae comprises 11 genera and includes salmon, trout, charr, freshwater whitefishes, ciscos and graylings [1]. Many salmonid species are of considerable economic, social and environmental importance. Salmonids contribute to local and global economies through aquaculture, wild stock fisheries and recreational sport fisheries. In addition, they are a traditional food source for aboriginal peoples in Canada and have a central role in their culture. Salmon and trout are sentinel species for monitoring the aquatic environment and therefore they are used extensively for ecotoxicology studies. As a result of human activities related to the rearing of salmon and trout and the need to make management decisions concerning stock assessment and harvesting plans, there is a large salmonid research community working on the biology, life histories, population dynamics, biogeography, phylogenetic relationships, physiology and nutrition of salmonids.

Some fundamental scientific questions can be explored using salmonid genomes. The common ancestor of salmon and trout experienced a whole genome duplication, and modern species may be considered pseudo-tetraploid as they are in the process of reverting to a stable diploid state [2]. This makes them ideal organisms for examining the consequences of genome and gene duplications, processes that are considered to have had

pivotal roles in generating gene diversity and the functional specialization found in modern vertebrates [3]. How a genome reorganizes itself to cope with duplicated chromosomes and the importance of gene duplications for evolution and adaptation are long-standing issues in biology that remain unresolved [3,4].

As illustrated in Table S1 in Additional file 1, no other group of fish species receives such comprehensive combined commercial and scientific attention as the salmonids [5], but as yet there is no genome sequence available for any salmonid. The genome of the Atlantic salmon (*Salmo salar*) was selected to be the reference sequence for all salmonids on the basis of its importance for the aquaculture industry and because so much previous work has been carried out at the genomic level on this species (Salmonid Genome Sequencing Workshop 2007, Simon Fraser University, Burnaby, BC). Having a robust salmonid reference sequence will allow next generation sequencing technologies to be used to obtain the sequences of other salmonids, such as rainbow trout and the Pacific salmon, more rapidly and at lower cost. An International Collaboration to Sequence the Atlantic Salmon Genome (ICSASG), representing researchers, funding agencies and industry from Canada, Chile and Norway, was formed to undertake this project. As suggested by the Toronto Statement on the pre-publication release of genomic data [6], in this open letter we provide details concerning: the approach adopted to sequence the Atlantic salmon genome; the plans for quality control of raw data, sequence assemblies, the integration of other genomic resources and annotation; and the types of analyses that are of particular interest to the ICSASG. It should be noted that the Toronto Statement [6] recognizes the right of the ICSASG to prepare the first large-scale description of the Atlantic salmon sequencing dataset and this open letter does not constitute that first publication.

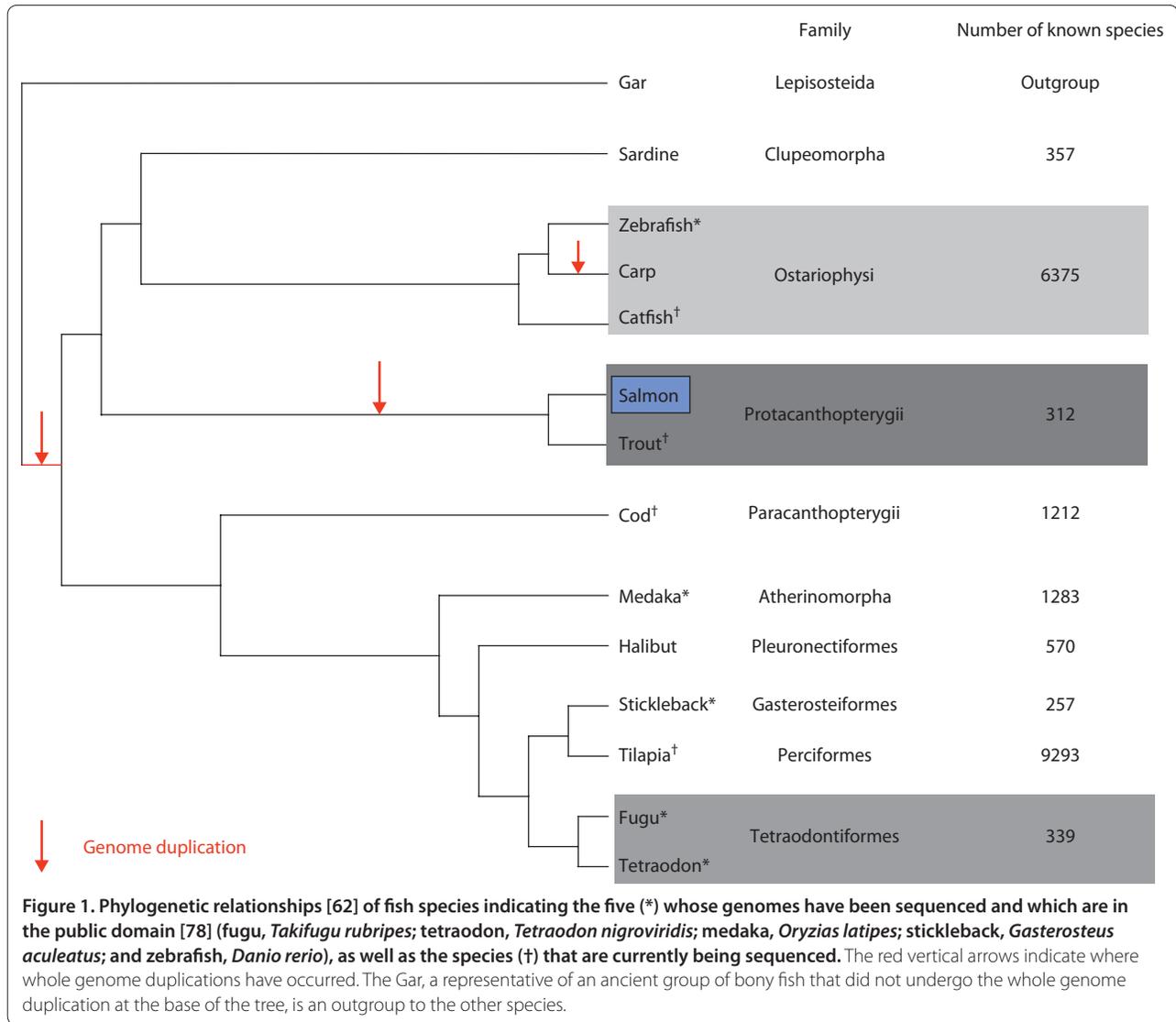
Position of salmonids in teleost phylogeny and their relevance for vertebrate genome evolution

Fish and mammals share a common ancestor that dates back to the origin of the vertebrate lineage. With the

*Correspondence: wdavidso@sfu.ca

¹Department of Molecular Biology and Biochemistry, Simon Fraser University, Burnaby BC, V5A 1S6, Canada

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



completion of the human genome sequence and several other mammalian genomes, it became clear that there is a need to investigate and characterize non-mammalian genomes to gain a more complete understanding of the complexity of vertebrate genomes and how they have evolved [7]. There are ~25,000 species of fish, making them the most successful vertebrate group, and their diversity makes them useful model systems in different disciplines of biology. The number of teleost species far exceeds those of any other fish group, or any other vertebrate, and this has been attributed to a whole genome duplication before their radiation in the Cretaceous Period [8-11]. In terms of phylogeny the salmonids hold a unique phylogenetic position compared with the fish species whose genomes have been sequenced or are in the process of being sequenced as they belong to the Protacanthopterygii, the most

primitive group of teleosts (Figure 1). Thus, the salmonids provide a key phylogenetic link between teleost evolution and the evolution of non-teleost fish as well as other vertebrates.

Properties of the Atlantic salmon genome and genomic resources currently available

Much of the basic information concerning the Atlantic salmon genome is known. For example, the C value for Atlantic salmon has been estimated as 3.27 pg [12], which translates into a haploid genome size of ~3 x 10⁹ bp. The G+C content of the Atlantic salmon genome is 44.4% [13]. Although the Atlantic salmon genome is fairly similar to those of warm-blooded vertebrates with respect to size and overall base composition, it seems to be devoid of isochores structures like other coldwater fish genomes [14]. This is reflected in the inability to obtain

G-banding patterns in Atlantic salmon chromosomes [15]. It has been suggested that the diploid ancestor of salmonids possessed a karyotype with 48 acrocentric chromosomes, resulting in 96 acrocentrics after the genome duplication [16]. Comparisons of the karyotypes of several salmonid species, including Atlantic salmon, revealed that many gross chromosomal rearrangements (fusions and inversions) have occurred along the different lineages since the ancestral whole genome duplication occurred [16]. The Atlantic salmon whose genome has been chosen to be sequenced represents the European subspecies (*S. salar europensis*), with 29 pairs of chromosomes [17,18].

Approximately 200 cDNA libraries have been constructed from many different tissues and developmental stages of Atlantic salmon [19-27]. As of 2 April 2010, there were 495,257 Atlantic salmon Expressed Sequence Tags (ESTs) [28]. Atlantic salmon ranked 20th by organism with respect to total number of ESTs, and almost all of the other organisms in the top 20 have had their genomes sequenced or are in progress of so doing. The ESTs have been placed in over 81,000 contigs and annotated [25]. As of 2 April 2010, 33,709 Atlantic salmon UniGenes had been identified [29] and over 9,057 reference quality full-length cDNA coding sequences had been confirmed [27]. All of this information can be viewed on publicly available websites [29-32]. The EST databases provide a rich source of material for identifying genetic markers, such as microsatellites [33,34] and single nucleotide polymorphisms (SNPs) [35,36], that have been used to place genes on linkage maps [37,38]. In addition, the ESTs enabled the construction of microarrays for expression analyses [22,25,39-41]. More than 60 groups around the world are using these microarrays, indicating that there is a large salmonid research community actively engaged in functional genomics. Moreover, the EST databases, especially the full-length cDNA coding sequences, will form the basis for building gene models during the annotation of the Atlantic salmon genome.

A publicly available Atlantic salmon bacterial artificial chromosome (BAC) library (CHORI-214) was constructed from the DNA of an individual male from a Norwegian aquaculture strain and arrayed on nylon membranes [42]. HindIII fingerprinting of the CHORI-214 Atlantic salmon BACs was used to create the first physical map of a salmonid genome, consisting of 223,781 BACs in ~4,565 contigs and 33,217 singletons [43]. 207,869 BAC end sequences with an average length of 666 bp (~3.5% of the genome) were produced to yield a snapshot of the Atlantic salmon genome and to identify putative syntenic relationships between the Atlantic salmon physical map and the fish genomes that have been sequenced ([44] and KA Boroevich, KP Lubieniecki, W Chow, P de Jong, J Schein, M Field, R Moore, JG de Boer, BFK, WSD,

unpublished results). Several linkage maps based on microsatellites, amplified fragment length polymorphisms (AFLPs) and SNPs have been constructed for Atlantic salmon [37,38,45-47]. The BAC end sequences also provide a rich source of microsatellite markers and SNPs, which were used to integrate the physical and linkage maps [37,47]. Fluorescent *in situ* hybridization (FISH) analysis with BACs that contain microsatellite markers map was used to assign chromosome arms to linkage groups [48]. All of these genomic resources have been made publicly accessible through a website ([49] and KA Boroevich, KP Lubieniecki, W Chow, P de Jong, J Schein, M Field, R Moore, JG de Boer, BFK, WSD, unpublished results).

An Illumina iSelect bead-array, designed to interrogate ~16,500 putative Atlantic Salmon SNPs, was developed at the Center for Integrative Genetics (CIGENE). Approximately 55% of the SNPs on the array were identified from EST alignments, with most of the remainder coming from a random genomic sampling following construction of reduced representation libraries produced from individual and pooled DNA samples and high-throughput 454 pyrosequencing (MP Kent, B Hayes, Q Xiang, PR Berg, RA Gibbs, S Lien, personal communication). The SNP array is currently being used to genotype a mapping population consisting of 3,500 individuals and construct a high-resolution SNP map for Atlantic salmon, which will be beneficial in assembling the Atlantic salmon genome sequence.

More than 60 Atlantic salmon BACs have been sequenced so far ([50-58], and see [59] for list). Along with the BAC end sequences these provide a snapshot of the organization of the Atlantic salmon genome. It is estimated that repetitive DNA accounts for 30-35% of the Atlantic salmon genome. Fourteen families of DNA transposons (twelve Tc1-like and two piggyBac-like) constituting 6-10% of the genome have been identified. These DNA transposons are approximately 1,500 bp in length, with different repeat families ranging in similarity from 80% to 94% [60]. An Atlantic salmon repeat database has been developed, and a salmonid repeat masking tool is publicly available [61].

The Atlantic salmon genome sequencing strategy

Although five fish genomes have been sequenced, they represent euteleostei lineages that have been separated from salmonids for at least 200 million years (Figure 1) [62]. The C values of these fish genomes range in size from 0.35 pg (spotted green pufferfish) to 1.80 pg (zebrafish) [63], whereas the Atlantic salmon genome is considerably larger (3.27 pg) [12]. The Atlantic salmon genome is further complicated by the autotetraploid whole genome duplication, which occurred ~25-100 million years ago in the common ancestor of extant

salmonids [2], and the ongoing re-diploidization process, which involves genome rearrangements and the loss, sub-functionalization and neo-functionalization of duplicated genes [3,4,64]. Comparisons of duplicated regions of the Atlantic salmon genome reveal that they are 81-89% identical (e.g. 81-85% identity over 225 kb in the IgH A and B regions [56] and 87-89% identity in the duplicate copies of growth hormone genes [52]). Similarly, EST contig duplicates show 91-93% identity [25,27]. Therefore, it should be possible to resolve and assemble complex duplicated regions of the Atlantic salmon genome. Nevertheless, the size and complexity of the Atlantic salmon genome, particularly the long and frequent repeats [60], combined with the lack of a closely related guide sequence, means that assembling the sequence of the Atlantic salmon genome will be challenging.

Some of the problems of assembling and characterizing duplicate segments of the genome will be overcome by the choice of the Atlantic salmon to be sequenced: a double-haploid individual that was produced by mitotic androgenesis. The fish was female and was nicknamed 'Sally'. The homozygous nature of Sally's genome was verified by screening for polymorphisms at ~70 micro-satellite loci. Karyotyping of Sally revealed that she had a haploid chromosome content of 29, in accordance of what would be expected for an Atlantic salmon from Norway. No apparent chromosomal rearrangements were observed; however, she does seem to have been a mosaic with ~30% haploid cells and the remainder diploid cells (U Grimholt, personal communication). The availability of DNA from a totally homozygous individual will greatly facilitate the assembly of whole genome shotgun sequences. Because the fish chosen for sequencing was female and male salmon are the heterogametic sex [65], the male sex-determining region will not be sequenced at this stage of the Atlantic salmon genome sequencing project.

The feasibility of using GS FLX pyrosequencing (shotgun and paired-end reads with an average read length of 250 bp) to sequence the Atlantic salmon genome was assessed by attempting to sequence six BACs that form a minimum tiling path estimated to be ~1 Mb [66]. The conclusion reached was that the GS FLX technology was limited to gene mining and establishing a set of ordered sequence contigs with many gaps. Although the pyrosequencing technology has been improved and average read lengths of 450 bp are now routinely achieved, this is still not sufficient to get at least halfway through a repeat of 1,500 bp, especially if paired end reads are desired. Indeed, the recent assembly of the cod genome seems to confirm this. A 27-fold coverage based on GS FLX Titanium reads (average read length of 336 bp), which included paired ends, produced contigs with an N_{50} of 2,400 bp and ~14,000 scaffolds with an N_{50}

of 571 kb that cover 618 Mb of the predicted 930 Mb genome (where N_{50} is defined as the contig length such that using equal or longer contigs produces half the bases of the genome). The N_{50} size is computed by sorting all contigs from largest to smallest and by determining the minimum set of contigs whose sizes total 50% of the entire genome. Currently, to obtain a genome sequence that can act as a reference sequence for other salmonids, it seemed that a substantial portion of the sequencing of the Atlantic salmon genome should be carried out using Sanger technology or an equivalent with respect to read length. The repetitive nature of the Atlantic salmon genome, and especially the length of the most common repeat (~1,500 bp) [60], make it necessary to have long paired end reads for assembling the sequence of this species' genome.

Description of sequencing project with anticipated milestones

The ICSASG has raised sufficient funds from the Research Council of Norway, the Norwegian Fishery and Aquaculture Industry Research Fund, Genome British Columbia, the Chilean Economic Development Agency and the InnovaChile Committee to cover the cost of sequencing, assembling and annotating the Atlantic salmon genome. In Phase 1, which began on 1 January 2010, Beckman Coulter Inc. will produce a fourfold coverage of the genome using paired end, plasmid, fosmid and BAC Sanger sequences by the end of January 2011. Phase 2 will be conducted using primarily next generation sequencing technologies and will result in a high definition, well-annotated genome. It is anticipated that Phase 2 will commence in January 2011 and that the sequencing component will be completed by September 2011. The assembled sequence will be integrated with the physical map, the linkage map and the ESTs. By the beginning of 2012 an automated pipeline annotation will be prepared and the sequence placed on a genome browser such as Ensembl [67]. The salmonid community will then be invited to participate in a manual annotation of the Atlantic salmon genome using the strategy advocated by Elsiek *et al.* [68]. We anticipate that the key paper and companion reports describing the Atlantic salmon genome and its insight into salmonid biology and vertebrate genome evolution will be published in the summer of 2012.

Biological questions and types of analyses to be undertaken by ICSASG

The availability of a complete genome sequence for Atlantic salmon will have a major impact on all sectors of the international salmonid community. For the aquaculture industry it will provide a complete suite of genetic markers for the identification of the genes and alleles

responsible for production traits (such as growth, disease resistance, feed efficiency and age of sexual maturation). Companies will be able to develop tailored broodstock more precisely using nucleotide- or allele-assisted selection rather than the more general marker-assisted selection. In conjunction with traditional breeding practices, this approach promises rapid gains that will make companies who embrace this technology more competitive. For those government agencies with a mandate to conserve and manage wild stocks, the sequence will provide the tools that will make it possible to identify and distinguish discrete populations of salmon using genes that are selected for specific environments (such as thermal tolerance) rather than neutral genetic markers. Thus, the genome sequence has great potential for enabling sensitive and more accurate assessments of the sustainability of salmonid populations. For the academic community, the Atlantic salmon genome will provide a reference sequence for other related genomes (such as rainbow trout, Pacific salmon or charr), which could be rapidly and cost effectively sequenced using novel sequencing technologies. This will allow comparative genomics to be incorporated into ecology and evolutionary and population biology and bring to the fore the concept of landscape evolutionary genomics [69]. For ecotoxicologists, the salmon sequence will permit a more robust use of salmonids as sentinel species for monitoring the quality of the aquatic environment using genes whose expression patterns are known to respond to particular pollutants.

Some specific questions and projects that researchers associated with the ICSASG are particularly interested in, and which provided the motivation for sequencing the Atlantic salmon genome, include: the characterization of the immune system in salmon [50,51,56-58] and how this relates to resistance to specific pathogens [44,70,71] that affect the aquaculture industry; the identification of the master sex-determining gene and the pathways that regulate sexual maturation [65,72-76]; an understanding of the biological cues and the sensing mechanisms that allow salmon to return faithfully to their natal streams after extensive marine migrations [53,54]; and the rediploidization process and the fates of duplicated genes after a whole genome duplication [25,27,52,55].

Formation of the ICSASG and a portal for salmonid genomic resources

Individual researchers were receiving grants from national agencies during the 1980s and 1990s to carry out genetic studies on salmonids, but the funding was not sufficient for any group to make significant headway on their own. The Norwegian Salmon Genome Project and the Canadian Genomics Research on Atlantic Salmon Project (GRASP), funded from 2001 until 2005, allowed great

progress to be made, but it was evident to the participants that even more could be achieved if they pooled resources and worked together. This was the basis for forming cGRASP: the Consortium for Genomics Research on All Salmonids Project, which was successful in bringing together salmonid genomics teams from Canada, Norway, Scotland and the USA.

In 2005 the need for a scientific organizational body to coordinate genome research efforts and ensure that existing and upcoming resources were made accessible worldwide was recognized by the international salmonid research community. At a workshop held on 25-26 October 2005 at the Norwegian University for Life Sciences in Ås, Norway, the Consortium for Genomic Research on All Salmonids Program (cGRASP) was formed. A follow-up meeting held on October 10-12, 2006 at Simon Fraser University, Burnaby, Canada, attracted representatives of the salmonid research community from 17 countries. These researchers identified very clearly that there was a need for at least one high quality, whole genome salmonid sequence to make optimal use of genomics tools within salmonid research, that many of the pre-sequencing phase resources would be in place for Atlantic salmon by the end of 2008 and that the research community was strongly committed to building and maintaining the necessary organizational apparatus for handling the pre-sequencing phase, the draft sequence phase and the initial annotation phase.

Given the benefit that the salmon genome sequence will bring to the aquaculture industry, government agencies charged with managing wild stocks and monitoring the aquatic environment, foundations concerned with conservation issues and the academic community at large, it was proposed that an international public-private partnership would be the most appropriate funding model to accomplish this task. At a meeting in Quebec City in 2008 representatives of funding agencies from British Columbia, Norway and Chile resolved to work together to sequence the Atlantic salmon genome, and in April 2009 in Santiago, Chile the ICSASG was formally established. The ICSASG has raised sufficient funds to cover the cost of sequencing, assembling and annotating the Atlantic salmon genome. It is anticipated that when the sequence becomes available, other opportunities will arise and the framework established by the ICSASG can be expanded to encompass projects such as sequencing other salmonid genomes using the Atlantic salmon as a reference sequence. After the meeting at Simon Fraser University in 2006 a website [77] was set up as a portal for other websites that host salmonid genomic data, as well as providing information concerning on-going projects, collaborative opportunities and contact information for the ICSASG.

Additional material

Additional file 1 Table S1. A table showing the number of publications since 1955 for fish species or groups of species.

Author details

¹Department of Molecular Biology and Biochemistry, Simon Fraser University, Burnaby BC, V5A 1S6, Canada. ²Centre for Biomedical Research, University of Victoria, Victoria BC, V8W 3N5, Canada. ³Genome Sciences Centre, BC Cancer Agency, Vancouver BC, V5Z 4E6, Canada. ⁴Programa de Genética Humana, ICBM, Facultad de Medicina, Universidad de Chile, Santiago, Chile. ⁵Departamento de Biología, Facultad de Química y Biología, Universidad de Santiago, Chile. ⁶Departamento de Ingeniería Matemática y Centro de Modelamiento Matemático, Universidad de Chile, Santiago, Chile. ⁷Department of Informatics, University of Bergen, Bergen, Norway. ⁸CIGENE, Norwegian University of Life Sciences, N-1432, Ås, Norway

Published: 29 September 2010

References

- Nelson JS: *Fishes of the World*. 3rd edition. New York: Wiley and Son; 2006.
- Allendorf FW, Thorgaard GH: **Tetraploidy and the evolution of salmonid fishes**. In *Evolutionary Genetics of Fishes*. Edited by Turner BJ. New York: Plenum Press; 1984:1-53.
- Ohno S: *Evolution by Gene Duplication*. New York: Springer; 1970.
- Wolfe KH: **Yesterday's polyploids and the mystery of diploidization**. *Nat Rev Genet* 2001, **2**:333-341.
- Thorgaard GH, Bailey GS, Williams D, Buhler DR, Kaattari SL, Ristow SS, Hansen JD, Winton JR, Bartholomew JL, Nagler JJ, Walsh PJ, Vijayan MM, Devlin RH, Hardy RW, Overturf KE, Young WP, Robison BD, Rexroad C, Palti Y: **Status and opportunities for genomics research with rainbow trout**. *Comp Biochem Physiol B Biochem Mol Biol* 2002, **133**:609-646.
- Toronto International Data Release Workshop Authors, Birney E, Hudson TJ, Green ED, Gunter C, Eddy S, Rogers J, Harris JR, Ehrlich SD, Apweiler R, Austin CP, Berglund L, Bobrow M, Bountra C, Brookes AJ, Cambon-Thomsen A, Carter NP, Chisholm RL, Contreras JL, Cooke RM, Crosby WL, Dewar K, Durbin R, Dyke SO, Ecker JR, El Emam K, Feuk L, Gabriel SB, Gallacher J, Gelbart WM, Granell A, et al.: **Prepublication data sharing**. *Nature* 2009, **461**:168-70.
- Genome 10K Community of Scientists: **Genome 10K: a proposal to obtain whole-genome sequence for 10,000 vertebrate species**. *J Hered* 2009, **100**:659-674.
- Jaillon O, Aury JM, Brunet F, Petit JL, Stange-Thomann N, Mauceli E, Bouneau L, Fischer C, Ozouf-Costaz C, Bernot A, Nicaud S, Jaffe D, Fisher S, Lutfalla G, Dossat C, Segurens B, Dasilva C, Salanoubat M, Levy M, Boudet N, Castellano S, Anthouard V, Jubin C, Castelli V, Katinka M, Vacherie B, Biemont C, Skalli Z, Cattolico L, Poulain J, et al.: **Genome duplication in the teleost fish *Tetraodon nigroviridis* reveals the early vertebrate proto-karyotype**. *Nature* 2004, **431**:946-957.
- Vandepoel K, De Vos W, Taylor JS, Meyer A, Van de Peer Y: **Major events in the genome evolution of vertebrates: paranome age and size differ considerably between ray-finned fishes and land vertebrates**. *Proc Natl Acad Sci USA* 2004, **101**:1638-1643.
- Meyer A, Van de Peer Y: **From 2R to 3R: evidence for a fish-specific genome duplication (FSGD)**. *BioEssays* 2005, **27**:937-945.
- Christoffels A, Koh EG, Chia JM, Brenner S, Aparicio S, Venkatesh B: **Fugu genome analysis provides evidence for a whole-genome duplication early during the evolution of ray-finned fishes**. *Mol Biol Evol* 2004, **21**:1146-1151.
- Hardie DC, Hebert PD: **The nucleotide effects of cellular DNA content in cartilaginous and ray-finned fishes**. *Genome* 2003, **46**:683-706.
- Bucciarelli G, Bernardi G, Bernardi G: **An ultracentrifugation analysis of two hundred fish genomes**. *Gene* 2002, **295**:153-162.
- Bernardi G: **The isochore organization of the human genome and its evolutionary history - a review**. *Gene* 1993, **135**:57-66.
- Hartley SE, Horne MT: **Chromosome relationships in the genus *Salmo***. *Chromosoma* 1984, **90**:229-237.
- Phillips R, Rab P: **Chromosome evolution in the Salmonidae (Pisces): an update**. *Biol Rev Camb Philos Soc* 2001, **76**:1-25.
- Hartley SE: **The chromosomes of salmonid fishes**. *Biol Rev* 1987, **62**:197-214.
- King T, Verspoor E, Spidle AP, Gross R, Phillips RB, Koljonen M-L, Sanchez JA, Morrison CL: **Biodiversity and population structure**. In *The Atlantic Salmon: Genetics, Conservation and Management*. Edited by Verspoor E, Stradmeyer L, Nielsen JL. Oxford: Blackwell; 2007:117-166.
- Davey GC, Caplice NC, Martin SA, Powell R: **A survey of genes in the Atlantic salmon (*Salmo salar*) as identified by expressed sequence tags**. *Gene* 2001, **263**:121-130.
- Martin SA, Caplice NC, Davey GC, Powell R: **EST-based identification of genes expressed in the liver of adult Atlantic salmon (*Salmo salar*)**. *Biochem Biophys Res Comm* 2002, **293**:578-585.
- Tsoi SC, Ewart KV, Penny S, Melville K, Liebscher RS, LL Brown LL, Douglas SE: **Identification of immune-relevant genes from Atlantic salmon using suppression subtractive hybridization**. *Mar Biotechnol* 2004, **6**:199-214.
- Rise ML, von Schalburg KR, Brown GD, Devlin RH, Mawer MA, Kuipers N, Busby M, M Beetz-Sargent M, Alberto R, Gibbs AR, Hunt P, Shukin R, Zeznik JA, Nelson C, Jones SRM, Smailus DE, Jones SJM, Schein JE, Marra MA, Butterfield YSN, Stott JM, Ng SH, Davidson WS, Koop BF: **Development and application of a salmonid EST database and cDNA microarray: data mining and interspecific hybridization characteristics**. *Genome Res* 2004, **14**:478-490.
- Hagen-Larsen H, Laerdahl JK, Panitz F, Adzhubei A, Hoyheim B: **An EST-based approach for identifying genes expressed in the intestine and gills of pre-smolt Atlantic salmon (*Salmo salar*)**. *BMC Genomics* 2005, **6**:171.
- Adzhubei AA, Vasova AV, Hagen-Larsen H, Ruden TA, Laerdahl JK, Hoyheim B: **Annotated expressed sequence tags (ESTs) from pre-smolt Atlantic salmon (*Salmo salar*) in a searchable data resource**. *BMC Genomics* 2007, **8**:209.
- Koop BF, von Schalburg KR, Leong J, Walker N, Lieph R, Cooper GA, Robb A, Beetz-Sargent M, Holt RA, Moore R, Brahmabhatt S, Rosner J, Rexroad CE 3rd, McGowan CR, Davidson WS: **A salmonid EST genomic study: genes, duplications, phylogeny and microarrays**. *BMC Genomics* 2008, **9**:545.
- Andreassen R, Lunner S, Hoyheim B: **Characterization of full-length sequenced cDNA inserts (FLICs) from Atlantic salmon (*Salmo salar*)**. *BMC Genomics* 2009, **10**:502.
- Leong JS, Jantzen SG, von Schalburg KR, Cooper GA, Messmer AM, Liao NY, Munro S, Moore R, Holt RA, Jones SJ, Davidson WS, Koop BF: ***Salmo salar* and *Esox lucius* full-length cDNA sequences reveal changes in evolutionary pressures on a post-tetraploidization genome**. *BMC Genomics* 2010, **11**:279.
- dbEST release 082710: Summary by Organism - August 27, 2010 [http://www.ncbi.nlm.nih.gov/dbEST/dbEST_summary.html]
- UniGene [http://www.ncbi.nlm.nih.gov/UniGene]
- Consortium for Genomic Research on All Salmon [http://web.uvic.ca/grasp/]
- The Gene Index Project: DFCI A.salmon Gene Index [http://compbio.dfci.harvard.edu/cgi-bin/tgi/gimain.pl?gudb=salmon]
- SalmonDB website [http://genomicasalmones.dim.uchile.cl]
- Ng SHS, Chang A, Brown GD, Koop BF, Davidson WS: **Type I microsatellite markers from Atlantic salmon (*Salmo salar*) expressed sequence tags**. *Mol Ecol Notes* 2005, **5**:762-766.
- Vasemagi A, Nilsson J, Primmer CR: **Expressed sequence tag-linked microsatellites as a source of gene-associated polymorphisms for detecting signatures of divergent selection in Atlantic salmon (*Salmo salar* L.)**. *Mol Biol Evol* 2005, **22**:1067-1076.
- Smith CT, Elfstrom CM, Seeb LW, Seeb JE: **Use of sequence data from rainbow trout and Atlantic salmon for SNP detection in Pacific salmon**. *Mol Ecol* 2005, **14**:4193-4203.
- Hayes B, Laerdahl JK, Lien S, Moen T, Berg P, Hindar K, Davidson WS, Koop BF, Adzhubei A, Hoyheim B: **An extensive resource of single nucleotide polymorphism markers associated with Atlantic salmon (*Salmo salar*) expressed sequences**. *Aquaculture* 2007, **265**:82-90.
- Danzmann RG, Davidson EA, Ferguson MM, Gharbi K, Koop BF, Hoyheim B, Lien S, Lubieniecki KP, Moghadam HK, Park J, Phillips RB, Davidson WS: **Distribution of ancestral proto-Actinopterygian chromosome arms within the genomes of 4R-derivative salmonid fishes (Rainbow trout and Atlantic salmon)**. *BMC Genomics* 2008, **9**:557.
- Moen T, Hayes B, Baranski M, Berg PR, Kjolglum S, Koop BF, Davidson WS, Omholt SW, Lien S: **A linkage map of the Atlantic salmon (*Salmo salar*) based on EST-derived SNP markers**. *BMC Genomics* 2008, **9**:223.
- von Schalburg KR, Rise ML, Cooper GA, Brown GD, Gibbs RA, Nelson CC, Davidson WS, Koop BF: **Fish and chips: various methodologies demonstrate utility of a 16,006-gene salmonid microarray**. *BMC Genomics* 2005, **6**:126.
- von Schalburg KR, Cooper GA, Leong J, Robb A, Lieph R, Rise ML, Davidson WS, Koop BF: **Expansion of the genomics research on Atlantic salmon**

- Salmo salar* L. project (GRASP) microarray tools. *J Fish Biol* 2008, **72**:2051-2070.
41. Taggart JB, Bron JE, Martin SA, Seear PJ, Høyheim B, Talbot R, Carmichael SN, Villeneuve LA, Sweeney GE, Houlihan DF, Secombes CJ, Tocher DR, Teale AJ: **A description of the origins, design and performance of the TRAITS-SGP Atlantic salmon *Salmo salar* L. cDNA microarray.** *J Fish Biol* 2008, **72**:2071-2094.
 42. Thorsen J, Zhu B, Frengen E, Osoegawa K, de Jong PJ, Koop BF, Davidson WS, Høyheim B: **A highly redundant BAC library of Atlantic salmon (*Salmo salar*): an important tool for salmon projects.** *BMC Genomics* 2005, **6**:50.
 43. Ng SHS, Artieri CG, Bosdet IE, Chiu R, Danzmann RG, Davidson WS, Ferguson MM, Fjell CD, Høyheim B, Jones SJM, de Jong PJ, Koop BF, Krzywinski MI, Lubieniecki K, Marra MA, Mitchell LA, Methewson C, Osegawa K, Parisotto SE, Phillips RB, Rise ML, von Schalburg KR, Schein JE, Shin H, Siddiqui A, Thorsen J, Wye N, Yanand G, Zhu B: **A physical map of the genome of Atlantic salmon, *Salmo salar*.** *Genomics* 2005, **86**:396-404.
 44. Li J, Boroevich KA, Koop BF, Davidson WS: **Comparative Genomics Identifies Candidate Genes for Infectious Salmon Anemia (ISA) Resistance in Atlantic Salmon (*Salmo salar*).** *Mar Biotechnol* 2010, doi:10.1007/s10126-010-9284-0.
 45. Moen T, Høyheim B, Munck H, Gomez-Raya L: **A linkage map of Atlantic salmon (*Salmo salar*) reveals an uncommonly large difference in recombination rate between the sexes.** *Animal Genet* 2004, **35**:81-92.
 46. Gilbey J, Verspoor E, McLay A, Houlihan D: **A microsatellite linkage map for Atlantic salmon (*Salmo salar*).** *Animal Genet* 2004, **35**:98-105.
 47. Lorenz S, Brenna-Hansen S, Moen T, Roseth A, Davidson WS, Omholt SW, Lien S: **BAC-based upgrading and physical integration of a genetic SNP map in Atlantic salmon.** *Anim Genet* 2010, **41**:48-54.
 48. Phillips RB, Keatley KA, Morasch MR, Ventura AB, Lubieniecki KP, Koop BF, Danzmann RG, Davidson WS: **Assignment of Atlantic salmon (*Salmo salar*) linkage groups to specific chromosomes: conservation of large syntenic blocks corresponding to whole chromosome arms in rainbow trout (*Oncorhynchus mykiss*).** *BMC Genet* 2009, **10**:46.
 49. Asalbase [<http://www.asalbase.org>]
 50. Yazawa R, Cooper GA, Hunt P, Beetz-Sargent M, Robb A, Conrad M, McKinnel L, So S, Jantzen S, Phillips RB, Davidson WS, Koop BF: **Striking antigen recognition diversity in the Atlantic salmon T-cell receptor alpha/delta locus.** *Dev Comp Immunol* 2008, **32**:204-212.
 51. Yazawa R, Cooper GA, Beetz-Sargent M, Robb A, McKinnel L, Davidson WS, Koop BF: **Functional adaptive diversity of the Atlantic salmon T-cell receptor gamma locus.** *Mol Immunol* 2008, **45**:2150-2157.
 52. von Schalburg KR, Yazawa R, de Boer J, Lubieniecki KP, Goh B, Straub CA, Beetz-Sargent MR, Robb A, Davidson WS, Devlin RH, Koop BF: **Isolation, characterization and comparison of Atlantic and Chinook salmon growth hormone 1 and 2.** *BMC Genomics* 2008, **9**:522.
 53. Johnstone KA, Lubieniecki KP, Chow W, Phillips RB, Koop BF, Davidson WS: **Genomic organization and characterization of two vomeronasal 1 receptor-like genes (*ora1* and *ora2*) in Atlantic salmon *Salmo salar*.** *Marine Genomics* 2008, **1**:23-31.
 54. Johnstone KA, Ciborowski KL, Lubieniecki KP, Chow W, Phillips RB, Koop BF, Jordan WC, Davidson WS: **Genomic organization and evolution of the vomeronasal type 2 receptor-like (OlfC) gene clusters in Atlantic salmon, *Salmo salar*.** *Mol Biol Evol* 2009, **26**:1117-1125.
 55. Lai YYY, Lubieniecki KP, Phillips RB, Chow W, Koop BF, Davidson WS: **Genomic organization of Atlantic salmon (*Salmo salar*) fatty acid binding protein (*fabp2*) genes reveals independent loss of duplicate loci in teleosts.** *Mar Genomics* 2009, **2**:193-200.
 56. Yasuike M, de Boer J, von Schalburg KR, Cooper GA, McKinnel L, Messmer A, So S, Davidson WS, Koop BF: **Evolution of duplicated *IgH* Loci in Atlantic salmon, *Salmo salar*.** *BMC Genomics* 2010, **11**:486.
 57. Lukacs MF, Harstad H, Grimholt U, Beetz-Sargent M, Cooper GA, Reid L, Bakke HG, Phillips RB, Miller KM, Davidson WS, Koop BF: **Genomic organization of duplicated major histocompatibility complex class I regions in Atlantic salmon (*Salmo salar*).** *BMC Genomics* 2007, **8**:251.
 58. Lukacs MF, Harstad H, Bakke HG, Beetz-Sargent M, McKinnel L, Lubieniecki KP, Koop BF, Grimholt U: **Comprehensive analysis of MHC class I genes from the U-, S- and Z- lineages in Atlantic salmon.** *BMC Genomics* 2010, **11**:154.
 59. **Genomic Research on Atlantic Salmon Project: BAC Sequences and Annotations** [<http://grasp.mbb.sfu.ca/bacannotations/GRASPBac.html>]
 60. de Boer JG, Yazawa R, Davidson WS, Koop BF: **Bursts and horizontal evolution of DNA transposons in the speciation of pseudotetraploid salmonids.** *BMC Genomics* 2007, **8**:422.
 61. **Centre for Biomedical Research at the University of Victoria: CBR RepeatMasker** [http://lucy.ceh.uvic.ca/repeatmasker/cbr_repeatmasker.py]
 62. Steinke D, Salzburger W, Meyer A: **Novel relationship among ten fish model species revealed based on a phylogeneomic analysis using ESTs.** *J Mol Evol* 2006, **62**:772-784.
 63. Gregory TR, Nicol JA, Tamm H, Kullman B, Kullman K, Leitch IJ, Murray BG, Kapraun DF, Greilhuber J, Bennett MD: **Eukaryotic genome size databases.** *Nucleic Acids Res* 2007, **35**:D332-D338.
 64. Force A, Lynch M, Pickett FB, Amores A, Yan Y, Postlethwait J: **Preservation of duplicate genes by complementary, degenerative mutations.** *Genetics* 1999, **151**:1531-1545.
 65. Davidson WS, Huang TK, Fujiki K, von Schalburg KR, Koop BF: **The sex determining loci and sex chromosomes in the family salmonidae.** *Sex Dev* 2009, **3**:78-87.
 66. Quinn NL, Levenkova N, Chow W, Bouffard P, Boroevich KA, Knight JR, Jarvie TP, Lubieniecki KP, Desany BA, Koop BF, Harkins TT, Davidson WS: **Assessing the feasibility of GS FLX pyrosequencing for sequencing the Atlantic salmon genome.** *BMC Genomics* 2008, **9**:404.
 67. Birney E, Andrews TD, Bevan P, Caccamo M, Chen Y, Clarke L, Coates G, Cuff J, Curwen V, Cutts T, Down T, Eyras E, Fernandez-Suarez XM, Gane P, Gibbins B, Gilbert J, Hammond M, Hotz HR, Iyer V, Jekosch K, Kahari A, Kasprzyk A, Keefe D, Keenan S, Lehvaslaiho H, McVicker G, Melsopp C, Meidl P, Mongin E, Pettett R, *et al.*: **An overview of Ensembl.** *Genome Res* 2004, **14**:925-928.
 68. Elsik CG, Worley KC, Zhang L, Milshina NV, Jiang H, Reese JT, Childs KL, Venkatraman A, Dickens CM, Weinstock GM, Gibbs RA: **Community annotation: procedures, protocols, and supporting tools.** *Genome Res* 2006, **16**:1329-1333.
 69. Lowry DB: **Landscape evolutionary genomics.** *Biol Lett* 2010, **6**:502-504.
 70. Rise ML, Jones SR, Brown GD, von Schalburg KR, Davidson WS, Koop BF: **Microarray analyses identify molecular biomarkers of Atlantic salmon macrophage and hematopoietic kidney response to *Piscirickettsia salmonis* infection.** *Physiol Genomics* 2004, **20**:21-35.
 71. Moen T, Baranski M, Sonesson AK, Kjøglum S: **Confirmation and fine-mapping of a major QTL for resistance to infectious pancreatic necrosis in Atlantic salmon (*Salmo salar*): population-level associations between markers and trait.** *BMC Genomics* 2009, **10**:368.
 72. Woram RA, Gharbi K, Sakamoto T, Høyheim B, Holm LE, Naish K, McGowan C, Ferguson MM, Phillips RB, Stein J, Guyomard R, Cairney M, Taggart JB, Powell R, Davidson W, Danzmann RG: **Comparative genome analysis of the primary sex-determining locus in salmonid fishes.** *Genome Res* 2003, **13**:272-280.
 73. Artieri CG, Mitchell LA, Ng SHS, Parisotto SE, Danzmann RG, Høyheim BJ, Phillips RB, Morasch M, Koop BF, Davidson WS: **Identification of the sex-determining locus of Atlantic salmon (*Salmo salar*) on chromosome 2.** *Cytogenet Genome Res* 2006, **112**:152-159.
 74. von Schalburg KR, Rise ML, Brown GD, Davidson WS, Koop BF: **A comprehensive survey of the genes involved in maturation and development of the rainbow trout ovary.** *Biol Reprod* 2005, **72**:687-699.
 75. von Schalburg KR, McCarthy SP, Rise ML, Hutson JC, Davidson WS, Koop BF: **Expression of morphogenic genes in mature ovarian and testicular tissues: potential stem-cell niche markers and patterning factors.** *Mol Reprod Dev* 2006, **73**:142-152.
 76. von Schalburg KR, Yasuike M, Davidson WS, Koop BF: **Regulation, expression and characterization of aromatase (*cyp19b1*) transcripts in ovary and testis of rainbow trout (*Oncorhynchus mykiss*).** *Comp Biochem Physiol B Biochem Mol Biol* 2010, **155**:118-125.
 77. **Consortium for Genomics Research on All Salmonids Program (cGRASP) portal** [<http://www.cgrasp.org>]
 78. **Ensembl Find a Species** [<http://www.ensembl.org/info/about/species.html>]

doi:10.1186/gb-2010-11-9-403

Cite this article as: Davidson WS, *et al.*: Sequencing the genome of the Atlantic salmon (*Salmo salar*). *Genome Biology* 2010, **11**:403.