The Salmon Genome Project
Project leader: Bjørn Høyheim

The Norwegian Salmon Genome Project was set up to construct a genetic map, a BAC library, cDNA libraries from various tissues, EST sequencing and a bioinformatics infrastructure as the major deliverables. All data and resources can be accessed through our web-site: www.salmongenome.no. Below is a brief overview of our available resources.

cDNA libraries and EST sequencing:
We have constructed 27 different libraries from 15 different tissues. A total of 160 000 clones have been picked and gridded from these libraries into 384 well microtiterplates and permanently stored. In addition all libraries have been amplified and permanently stored.
From these libraries we have sequenced nearly 68 000 ESTs. All EST sequences have been processed using our internal preprocessing pipeline to perform quality control and vector clipping, marking repeats and poly-A tails. To avoid data that might interfere with the analysis only sequences considered to be OK after the preprocessing was used in further analysis.
A total of 57 212 sequences have been clustered and subjected to automatic annotation based on Gene Ontology. The result after clustering was over 6 200 contigs and over 13 300 single sequences. Possible function has been deduced for a number of sequences. The results are sorted according to Molecular Function, Biological Process or Cellular Component.
A total of 55 118 processed sequences excluding mitochondrial sequences has been submitted to GenBank.

Full-length cDNAs:
A total of 1 100 cDNAs have been sequenced in full to give a complete sequence of the coding region of the genes.

Microarray chip:
A microarray chip has been developed consisting of nearly 17 000 different cDNAs in duplicate. The chip has been developed in collaboration with groups in the UK TRAITS consortium (www.abdn.ac.uk/sfirc/salmon/) and consists of approx. 60% clones from SGP, 30% from the EU funded SALGENE and 10% from TRAITS.
We have also constructed a smaller chip consisting of nearly 5 000 different cDNAs that we have used in studying developmental issues.

Genetic map:
A genetic map has been constructed. A total of 1 249 microsatellite sequences has been developed and submitted to GenBank. More than 600 markers has been genotyped in our resource families and a framework map consisting of over 400 markers has been constructed. In addition we have performed a clustering of all SGP and Canadian (GRASP) sequences to search for SNPs. We preprocessed all the raw data submitted by the Canadians to exclude bad data that might interfere with the analysis. The good sequences included approx. 104 000 EST sequences from GRASP and SGP and resulted in 14 958 contigs and 21 261 single sequences. The SNP search resulted in identification of nearly 2 500 SNPs.

BAC-library:
An Atlantic salmon (*Salmo salar*) BAC library has been constructed from DNA prepared from sperm of a single male by PhD student Jim Thorsen from NVH (CHORI 214). The library has been constructed in collaboration with Dr. Pieter de Jong at BACPAC Resources in Children's Hospital Oakland Research Institute and the Canadian project Genetic Research on Atlantic salmon Project (GRASP). The library consists of 313 000 clones, representing 18 fold genome coverage, which have been arrayed into 384-well microtiter dishes. The clones in the library have an average insert size ranging from 170 kb to 197 kb with a none-insert rate of 2-3%. The library is printed in duplicate on filters for screening.
Together with the Canadian GRASP project (web.uvic.ca/cbr/grasp/) we have participated in developing a physical map of these BACs. A total of 200 000 BACs has been fingerprinted and a first generation physical map has been constructed.

Bioinformatics infrastructure:
Three computers to store and handle all the data accumulated are installed at USIT, University of Oslo. This includes both a central database as well as all the necessary software to analyse the data.
Software have both been developed and installed on several computers in Bergen, Oslo and Trondheim for modelling, searching, sequence analysis, expression analysis, statistics and other relevant local software.
The web-based SGP data management system has two major components: a relational database and a web site www.salmongenome.no which contains both general information and tools that can be accessed by all groups interested. Database access is performed through the data search and retrieval interface. SGP maintains public services available on the project web site: the Blast server and pre-assembly sequence processing pipeline utilizing STADEN package. The latest versions of Blast databases are maintained and made available to other groups. The sequence processing pipeline can also be downloaded and used locally. SGP public services are also available via the FUGE bioinformatics platform at www.bioinfo.no.
The data management system also incorporates protected data and software that is only available to registered users after log in. Users can access more tools currently only open internally to the SGP partners. This includes a better BLAST service, a sequence analysis pipeline directly linked to the SGP database and a genetic map site. In addition to this two data processing pipelines had been developed and installed on a recently upgraded Magnum supercomputer at USIT. One of the pipelines allows to perform clustering of the sequencing data into contigs and SNP identification, the other pipeline is used for sequence annotation and the assignment of GO (Gene Ontology Consortium) definitions. Both pipelines have been extensively used for the SGP data processing.

Other research:
The Salmon Genome Project has performed research into the organisation of the salmon genome, to identify genes that are involved in disease resistance and in biodiversity studies to characterise farmed and wild stocks of Atlantic salmon.
The intron/exon boundaries has been studied using a selected set of nearly 400 ESTs that we constructed primers from to sequence the corresponding genomic sequence.
Functional studies have begun using Representational Differential Analysis (RDA) and macro- and microarray to investigate change in gene expression in salmon challenged with furunculosis or Infectious Salmon Anemia (ISA) compared to untreated salmon.