Ssa miRNAs DB: Online repository of *in silico* predicted miRNAs in *Salmo salar*

Daniela Reyes, Victoria Cepeda, Ruth González & Rodrigo Vidal*

Laboratory of Molecular Ecology, Genomics and Evolutionary Studies, Department of Biology, Faculty of Chemistry and Biology, Santiago de Chile University, Chile & Centro de Biotecnología Acuícola (CBA); Rodrigo Vidal – Email: ruben.vidal@usach.cl; Phone: +56 (02)7181120; Fax: +56 (025) 7181234; *Corresponding author

Abstract:
The Atlantic salmon (*Salmo salar*) is a very valuable commercial salmonid species. As with other aquaculture species, intensive aquaculture of Atlantic salmon often faces disease problems especially in early life stages which can limit stable production of the species. 'Ssa miRNAs DB', a bioinformatics and manually curated database, aims at providing a comprehensive resource of microRNA in Atlantic salmon, with a user friendly interface for a convenient retrieval of each entry by microRNA ID or target gene. The current version of Ssa miRNAs DB involved the prediction of 41 and 266 homologous and novel microRNAs, respectively.

Availability: http://www.molgenv.com/ssa_mirnas_db_home.php.

Keywords: Atlantic salmon, microRNA, HTML, Database

Background:
The Atlantic salmon (*Salmo salar*) is one of the most valuable commercial salmonid species with well-established markets in Europe, Japan and the United States, being the major producers Norway, Chile, the United Kingdom, and Canada. One of the major problems related to the mass production of salmonids, is the development of diseases, affecting directly survival and production rates. To efficiently manage disease problems in Atlantic salmon aquaculture is necessary improve our knowledge of the transcriptome and their regulation on this specie. One of the most important regulators of the gene expression is microRNAs (miRNAs). miRNAs are small, highly conserved, single-stranded RNAs generated from hairpin-shaped transcripts (pre-miRNA) **[1],** which regulates gene expression at post-transcriptional level by translational repression or cleavage of the target mRNA, being involved in several cellular processes, such as diseases and virus induced immune response **[2]**. In this work we develop "Ssa miRNAs DB", an online repository of in silico predicted *Salmo salar*

miRNAs and their targets. The identification of miRNAs in Atlantic salmon sequences was performed following an strategy based on the most distinctive characteristics of miRNAs: a) high conservation and size, b) pre-miRNA secondary structure and energetic stability, and c) complementarity with miRNA*.

Methodology:
Data collection
Transcriptomic sequences of Atlantic salmon from Unigene database of GenBank (http://www.ncbi.nlm.nih.gov/unigene), Gene Index (http://compbio.dfci.harvard.edu/tgi/), and GRASP (http://web.uvic.ca/grasp/) were downloaded, as well all sequences of mature miRNAs of vertebrates deposited on miRBase, release 16, and (http://www.mirbase.org) **[3]**.

In silico prediction of Atlantic salmon miRNAs
To identify miRNAs on sequences of Atlantic salmon a local alignment was performed between vertebrate miRNAs and salmon sequences using BLASTN **[4]** (word length=7, low

BIOINFORMATION

complexity filter=off), choosing only gapless, plus/plus alignments with a maximum of 3 mismatches. Fragments of ~420 nt were generated from selected hits (200 nt in both directions from the alignment).

Secondary structure of each fragment was predicted with RNAfold [5] and then analyzed with an in house developed script to identify hairpin-shaped structures (minimal stem length = 13 nt). Minimal Folding Free Energy Index (MFEI) was calculated with RNAfold and in house developed script for every hairpin detected, discarding all hairpins with an MFEI lesser than 0.85. All remaining hairpins were analyzed with miPred [6], determining if is pre-miRNA-like hairpin or not.

Figure 1: A snap shot of Ssa miRNAs DB.

BIOINFORMATION

Classification and annotation

The pre-miRNA-like hairpins where classified, using an in house developed script, as pre-miRNA of an homolog miRNA or precursors of a novel miRNA, based in the information of the local alignment, as the location of the mature miRNA and complementarity with miRNA*. For precursors of novel miRNAs the location and sequence of the mature miRNA was determined by MatureBayes **[7].** All the precursors and mature miRNAs were annotated according to the standard set by miRBase. miRNA target prediction. The target prediction of the identified miRNAs was performed using RNAhybrid **[8]** and miRanda **[9],** with settings that allows seek perfect complementarity of the seed region of the mature miRNA. Only those targets predicted by both programs were selected, to diminish the rate of false-positive results.

Utility:

miRNAs are small molecules, which play a crucial role in gene regulation. Several functional genomics efforts are currently in progress in salmonids species to help deciphering their function in immunological and biological basic processes. We believe that a detailed knowledge of the relationship between miRNA and their genomic environment provides extremely valuable information towards a better understanding of their functional role. In this context, was developed Ssa miRNAs DB, a database that provides an easy-to-use web interface retrieval of miRNA genomic data. A total of 41 homologous miRNAs and 266 novel miRNAs were identified in Atlantic salmon, which have approximately 30 thousand targets predicted by RNAhybrid and miRanda simultaneously. All the information about predicted miRNAs and their targets have been stored in a relational database developed in mySQL, with a PHP web interface hosted in http://www.molgenv.com/ssa_mirnas_db_home.php **(Figure 1).**

References:

[1] Bartel DP *Cell*. 2004 **116**: 281 [PMID: 14744438]
[2] Xiao C & Rajewsky K, *Cell*. 2009 **136**: 26 [PMID: 19135886]
[3] Griffiths Jones S *et al. Nucleic Acids Res*. 2006 **34**: D140 [PMID: 16381832]
[4] Altschul SF *et al. J Mol Biol.* 1990 **215**: 403 [PMID: 2231712]
[5] Zuker M & Stiegler P, *Nucleic Acids Res*. 1981 **9**: 133 [PMID: 6163133]
[6] Jiang P *et al. Nucleic Acids Res*. 2007 **35**: W339 [PMID: 17553836]
[7] Gkirtzou K *et al. PLoS One*. 2010 **5**: e11843 [PMID: 20700506]
[8] Rehmsmeier M *et al. RNA.* 2004 **10**: 1507 [PMID: 15383676]
[9] Enright AJ *et al. Genome Biol*. 2003 **5:** R1 [PMID: 14709173]