NORTH ATLANTIC SALMON WHOLE GENOME SEQUENCING ANALYSIS

A. Z. Pedrosa[1], B. D. Kovács[2], C. B. Sequeros[1]
[1]DTU Bioinformatics, [2]DTU Aqua, Technical University of Denmark

Introduction and aim

Selective breeding programs for Atlantic salmon (*Salmo salar*) are essential in aquaculture for improving economically important traits. Recently, the development of sequencing technologies and the generation of massive amounts of genomic data has allowed breeders to gain a deeper understanding of the genetic basis of these traits and thus helped breeding efforts. Resistance to a disease like **infectious salmon anemia** (ISA), a viral disease that generates big losses every year, is an interesting target in breeding and was studied in this project.

Following the discoveries of *Holborn et al (2020)*[1], who found 3 SNPs that explained over 20% of the ISA resistance variability in a Saint John River population of salmon, and of *LeBlanc et al (2010)*[2], who described a series of genes that are differentially expressed in fish with this disease, our aims are to:

I. Evaluate the structure of the populations under study with multidimensional scaling and identity by descent.
II. Search for the previously described ISA resistance-associated SNPs.
III. Look for high Fst regions between the two stocks and check whether they include any genes related to ISA.

Materials

Sequencing data generated by *Gao et al (2020)*[3] on a total of 7 individual samples from two North American Atlantic Salmon stocks propagated at the NCWMAC were analyzed: 4 from St. John River, of great economic importance, and 3 from Penobscot River, an endangered population.

We focus on chromosomes ssa11, ssa12 and ss13 for our studies and use the ICSASG_v2 reference genome[5].

Saint John River (SJR)	Penobscot River (PR)
SRR9925703[4] SRR99257044 SRR9925706[4] SRR9925707[4]	SRR9925722[4] SRR9925723[4] SRR9925724[4]

Methods

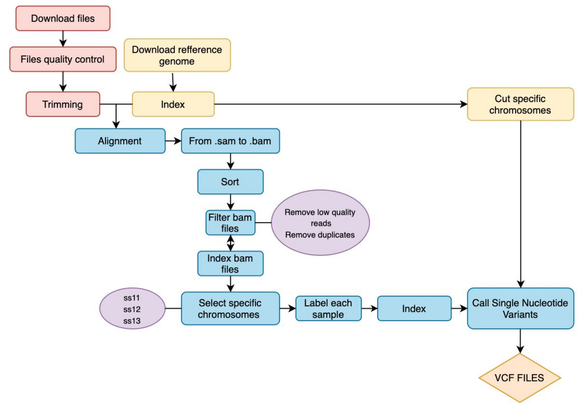

Figure 1. Data pre-processing, alignment and SNP calling on individual samples

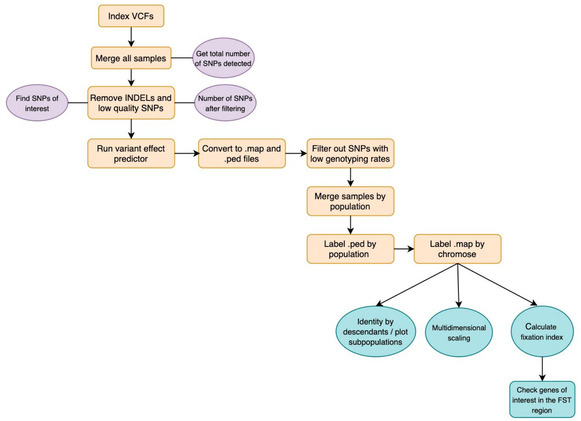

Figure 2. SNP analysis.

Results

After processing the raw NGS reads, we improved their quality significantly (Fig. 3). The alignments reached coverages of between 11x and 20x, with between 200 and 400 million reads per sample.

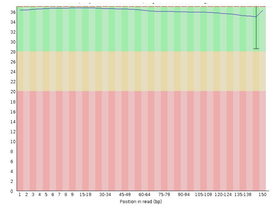

Figure 3. Per base sequence quality after processing

Looking at the population structure with multidimensional scaling, we see that the two are not very well differentiated in the studied chromosomes, which could be due to the small sample size. The identity by descent analysis shows similar results, with almost equal P-HATs in all sample comparisons.

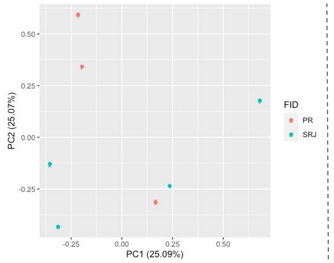

Figure 3. Multidimensional scaling. Red dots represent PR samples; blue dots, SRJ samples.

SRJ	03	SRJ	06	0.7760
SRJ	06	PR	22	0.7770
SRJ	03	PR	24	0.7784
SRJ	04	PR	22	0.7817
SRJ	04	PR	24	0.7819
PR	22	PR	23	0.7886
SRJ	03	PR	22	0.7892
SRJ	03	SRJ	04	0.7902
PR	22	PR	24	0.8011
PR	23	PR	24	0.8049

Figure 4. Identity By Descent analysis.

None of the SNPs shown to be associated with ISA resistance in *Holborn et al. (2020)*[1] were present in our samples. We attribute this to the small sample size or to a slight difference in the population under study. However, some genes differentially expressed in ISA-infected fish[2] were found to be inside of the higher Fst windows (Fig. 6), which could hint at a genetic difference in the two populations.

Some of these genes code for the ferritin heavy subunit in ssa11, a Ras-association domain-containing protein in ssa12 and a benzodiazepine receptor in ssa12.

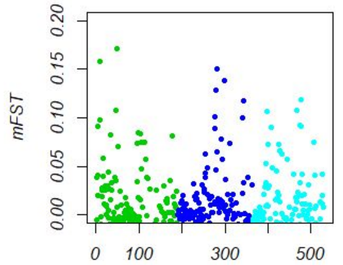

Figure 6. Fixation index in chromosomes ssa11 (green), ssa12 (dark blue) and ssa13 (light blue). The Fst is calculated in windows of 500kb.

Conclusions

Even though we cannot find the specific SNPs previously described to be associated with resistance to ISA in the Saint John River salmon, the fact that we see some ISA-related genes in high Fst regions of chromosomes ssa11, 12 and 13 could mean that there are some genetic differences in the two populations regarding resistance to this disease. The population structure analysis with MDS and IBD is not particularly informative, as we are working with a small number of samples. For future studies, more samples, possibly more populations and also a better understanding of the phenotypes in those populations would be of the essence to better understand the genetics of ISA.

References

[1]Genome wide analysis of infectious salmon anemia resistance in commercial Saint John River Atlantic salmon (2020), M. K. Holborn, K. P. Ang, J.A.K. Elliott, F. Powell, E. G.
[2]Genetic markers of the immune response of Atlantic salmon (Salmo salar) to infectious salmon anemia virus (ISAV) (2010), F. LeBlanc, M. Laflamme, N. Gagné
[3]A New Single Nucleotide Polymorphism Database for North American Atlantic Salmon Generated Through Whole Genome Resequencing (2020), G. Gao, M. R. Pietrak, G. S. Burr, C. E. Rexroad, B. C. Peterson, Y.
[4]https://www.ncbi.nlm.nih.gov/sra?LinkName=bioproject_sra_all&from_uid=559280
[5]https://www.ncbi.nlm.nih.gov/assembly/GCF_000233375.1

Technical University of Denmark