A genome-wide association study for resistance to viral nervous necrosis in Atlantic cod using a 12K single nucleotide polymorphism array

***R. Bangera*[*], M. Baranski[*], S. Lien[†]**

[*]Norwegian Institute of Food, Fisheries and Aquaculture Research, Tromsø, Norway, [†]Department of Animal and Aquacultural Sciences and Centre for Integrative Genetics, Norwegian University of Life Sciences, Ås, Norway

ABSTRACT: A genome-wide association study (GWAS) was performed for viral nervous necrosis (VNN) resistance in Atlantic cod using a polygenic (linear) mixed model approach. 704 offspring from 66 half-sib and 75 full-sib families in the Norwegian Atlantic cod breeding program were challenged tested under controlled conditions with nodavirus. The two disease traits; binary survival (BS) and number of days alive (ND) were analyzed using 8358 SNPs distributed across the Atlantic cod genome. In total, 29 and 36 genome-wide significant SNPs (P <0.0022) were detected for BS and ND, respectively. The heritability (h^2) estimated using genomic kinship matrix calculated from all SNPs was high for both BS (0.49) and ND (0.81). Linkage analysis and GWAS using alternative models ar also being performed to confirm the associations discovered, and these markers have potential for applied use in MAS or genomic selection schemes.

Keywords: Quantitative trait loci; Genome-wide association study; Polygenic model; Disease

Introduction

Aquaculture has always involved a high level of risk in terms of disease and spreading of infectious pathogens. The added stress from biological, physical and chemical factors may create an imbalance in the interaction between fish and pathogens leads to disease problems, which may be otherwise harmless under natural (wild) conditions (Wedemeyer, 1996) Genetic selection for disease resistance in farmed fish is generally based on survival after infection with a specific pathogen in a controlled challenge test for full-sib family groups from the breeding nucleus (Ødegård, et al., 2011). However, the major disadvantage of this method is the inability to record disease resistance on selection candidates due to disease transmission risks, which limits the accuracy of selection for such traits.

Advancement in genomics has enabled individual selection for disease resistance based on molecular information. Genetic markers linked to disease resistance can be used to improve the accuracy of selection through Marker-Assisted Selection (MAS). Recently, in Atlantic salmon, a major quantitative trait locus (QTL) affecting resistance to IPN (Houston, et al., 2008; Moen, et al., 2009) has been identified and successfully applied in MAS in both Norwegian and Scottish populations. An alternative approach for more polygenic traits is genomic selection, where genetic markers covering the whole genome are used so that all QTL are in linkage disequilibrium (LD) with at least one marker and selection is based on genetic values predicted from all the markers (Goddard, et al., 2007; Hayes, et al., 2001). The availability of high density single nucleotide polymorphism (SNP) arrays in livestock and now increasingly in aquaculture species has facilitated both genomic selection and genome-wide association studies (GWAS).

Viral nervous necrosis (VNN), caused by nodavirus, is a serious threat in aquaculture of marine species including Atlantic cod (*Gadus morhua* L.). Previous studies showed high heritability estimates (>0.68) for VNN resistance in Atlantic cod (Bangera, et al., 2011) indicating good prospects for genetic selection. In addition, a QTL study using microsatellite markers detected five genome-wide significant QTL for resistance to VNN, explaining a total of 68% of the phenotypic variance for survival (Baranski, et al., 2010). A 12K SNP array for Atlantic cod has recently been developed containing markers distributed across all 23 chromosomes, and is an excellent resource for fine-mapping and verification of the QTL detected in the earlier study.

The aim of this study was to validate and fine-map QTL for VNN resistance in a mapping population of Atlantic cod derived from the national cod breeding program in Norway.

Materials and Methods

The data used in this study originated from National Atlantic cod breeding program run by Nofima (Tromsø, Norway).

Fish material and traits. A total of 1471 juveniles (belonging to 152 families; produced from 110 sires and 152 dams) from year-class 2009 (second generation) were challenge tested with nodavirus by intramuscular injection as described in Ødegård et al. (Ødegård, et al., 2010)Mortality was recorded on a daily basis until the end of the test period which lasted for 35 days. The survival data included phenotypic trait recorded either as binary survival (dead or alive; BS) or number of days fish survived (ND). The overall survival at the end of the test was 29.50 %. The tissue samples for DNA isolation was collected from all 1471 fish and stored in 95% ethanol.

SNP genotyping. Genomic DNA was extracted from 704 offspring and their parents (66 sires and 75 dams) (average 10 offspring per full-sib family) Genotyping was performed with an Illumina Atlantic cod 12K SNP array (manuscript in preparation) at CIGENE in Norway

(www.cigene.no). Quality control of genotype data was performed and SNPs with minor allele frequency lower than 2.5% and those deviating from Mendelian segregation rules were discarded, resulting in a total of 8358 SNPs distributed across the genome that were used for further analysis.

Linkage map. The SNPs included in the study was incorporated in a linkage map covering 23 linkage groups (LG) and representing the expected 23 chromosomes of Atlantic cod map (manuscript in preparation).

Genome-wide association mapping. A polygenic model was used for GWAS for both survival traits (BS and ND) in order to identify loci associated with susceptibility and endurance to VNN resistance. This is a liner mixed model approach in which phenotype is regressed over SNP effects (as fixed) and polygenic effects (as random). A family based association using polygenic model make use of Genome-wide Rapid Analysis using Mixed Models and Score test (GRAMMAS) (Amin, et al., 2007) implemented in the GenABEL software package (Aulchenko, et al., 2007). Both traits were treated as quantitative trait, and genomic kinship coefficient estimated from SNP data was used in the polygenic model and maximum likelihood estimates (MLEs) are obtained using the available data. GRAMMAR uses these MLEs to estimate environmental residuals to run a simple score test which follows χ^2_{1df} and nominal P-values are estimated. Genome-wide association results were plotted using Haploview (Broad Institute, MIT, USA) with $-\log_{10}$ of nominal P-value and across chromosomal positions of each locus.

Significance testing. A Bonferroni correction ($\alpha^* = \alpha/n$) was applied with a predefined family wise error rate (FWER), $\alpha = 0.05$, to determine the genome-wide significant threshold p-values. A chromosome-wide threshold was adjusted for the number chromosomes in Atlantic cod (n=23). At 5% threshold, a SNP was considered genome-wide significant when chromosome-wide significant nominal (asymptotic) P-value was less than 0.0022 (i.e. $\alpha^* = 0.05/23$) or $-\log_{10}$ (P-value) of 2.66.

Results and Discussion

The GWAS analysis using for VNN resistance revealed 29 and 36 genome-wide significant SNPs ($P < 0.0022$) for BD and NS traits respectively (Figure 1 and Table 1). The 29 significant BS SNPs were found on all linkage groups with the exception of LG 9, 13, 15, 19 and 21 and the 36 significant ND SNPs were found on all linkage groups with the exception of LG 14, 19, 20 and 21. Among these significant SNPs, 12 were found to be common across traits. Moreover, 8 and 21 SNPs were found to be highly significant ($P < 0.00043$, at $\alpha = 0.01$) for BS and ND, respectively and of the top 10 SNPs, three were common across traits and mapped to LG 2 (at 27cM), 5 (at 64.7cM) and 23 (at 0cM). The Q-Q plot of the nominal P-values also suggests strong association between some of the genotyped SNPs and both the traits, given the deviation of the observed association statistics from the association statistics expected under null hypothesis of no association (Figure 2.).

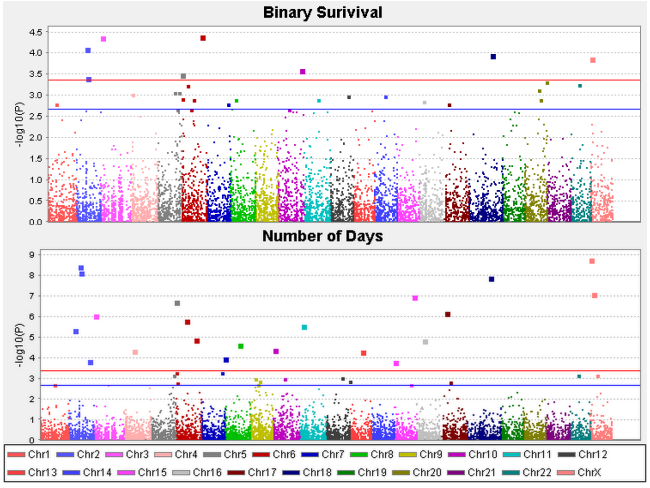

Figure 1. Results from genome-wide association studies with 8358 genome-wide SNPs are presented as –log(P) values for traits binary survival and number of days alive. Adjusted -log10(P) values for thresholds of 2.66 (lower line) for α =0.05 and 3.36 for α =0.01 (upper line)

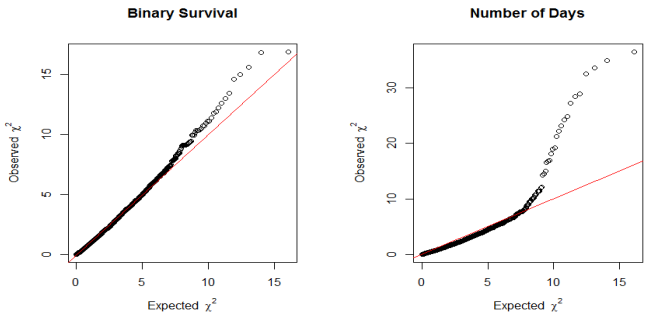

Figure 2. The quantile-quantile (Q-Q) plot shows the expected distribution (solid line) and the observed test statistics plotted against the expected test statistics (black dots)

Twelve significant SNPs were common across both traits, indicating that 'endurance' and 'susceptibility' have a degree of common genetic basis. However, susceptibility and endurance have been shown to only have a moderate genetic correlation (0.47) for VNN and suggests these two are in fact separate traits (Bangera, et al., 2013). In an earlier study, highly significant QTLs for VNN resistance have been mapped to LG5, LG9, LG10, LG20 and LG23(tentative linkage group designations based on current unpublished linkage map) (Baranski, et al., 2010). Given the marker density, the SNPs identified in this study are likely to be in LD with the causal variants rather than being the causal variants themselves (Bush, et al., 2012). Nevertheless, it may be possible to identify candidate genes through additional sequencing and use of the Atlantic cod genome sequence reference (Grisart, et al., 2002).

Table 1. Ten highest scoring significant SNPs based on genome-wide association study, for VNN resistance defined as Binary survival (BS) and Number of days alive (ND), are given. Linkage group (LG), sex averaged linkage map position (Pos) and nominal P-value.

Trait	LG	Pos (cM)	P-value	SNP
BS	6	50.6	0.00004	NS:172051_518
	3	2.6	0.00004	Gdist:310423_2194
	2	27	0.00008	Gdist:287668_1639*
	18	59.2	0.00011	NS:141159_861
	23	0	0.00013	Gdist:670512_2173*
	10	61.5	0.00025	NS:47013_1180
	5	64.7	0.00031	Gdist:351958_2891*
	2	28.9	0.00038	Gdist:68419_437
	20	56.8	0.00048	Gdist:188414_3998
	22	22	0.00057	Gdist:78083_252
ND	23	0	0.000000	Gdist:670512_2173*
	2	27	0.000000	Gdist:287668_1639*
	2	28.9	0.000000	Gdist:68419_437
	18	59.2	0.000000	NS:141159_861
	23	5.7	0.000000	NS:422145_459
	15	49.9	0.000000	Gdist:182434_1758
	5	64.7	0.000000	Gdist:351958_2891*
	17	11.5	0.000001	Gdist:97647_195
	3	2.6	0.000001	Gdist:310423_2194
	6	25	0.000002	Gdist:330183_823

*SNPs common in both traits

GWAS analyses for both traits were conducted using polygenic mixed model approach where kinship between individuals was modelled from genomic relationship matrix. Use of genomic kinship matrix instead of pedigree kinship is preferred for higher power of QTL detection, especially when trait with high heritability is considered. (Aulchenko, et al., 2009). Moreover, genomic-based kinship reflects true genetic relationship between the individuals than pedigree based relationships.

The heritability (h^2) as an estimate of the proportion of phenotypic variance that is explained by genomic relationship matrix (additive genetic variance) calculated from all 8358 genome-wide distributed SNPs was high when survival trait treated either as binary (0.49) or number of days alive (0.81). These estimates are in agreement with our earlier h^2 estimates from the same families (Bangera, et al., 2011; Bangera, et al., 2013). Significance testing in polygenic model using GRAMMAS expected to have high power of QTL detection when the trait is having high heritability (Aulchenko, et al., 2009).

Conclusion

The SNP associations identified in this study have considerable potential for applied use in selective breeding through MAS or genomic selection. However, further study on the extent and pattern of LD across the Atlantic cod genome is needed before the optimal MAS or genomic selection strategy can be chosen. The numerous significant SNPs distributed across a large number of chromosomes for both traits suggests large effect loci are present but that overall the high heritability for survival and endurance to VNN in cod is of a polygenic nature. Further analyses using linkage and alternative GWAS models are being undertaken in order to further characterize the most important genomic regions.

Acknowledgments

We thank all colleagues at the National cod breeding center (Nofima, Tromsø) and Fish Health Unit of Tromsø Aquaculture Research station. Ministry of Fisheries and Coastal Affairs, Norway will be highly acknowledged for their full financial support for the breeding program.

Literature Cited

Amin, N., van Duijn, C.M., Aulchenko, Y.S., 2007. PloS one. 2, e1274.

Aulchenko, Y., Struchalin, M., Aulchenko, M.Y., 2009.

Aulchenko, Y.S., Ripke, S., Isaacs, A., et al., 2007. Bioinformatics. 23, 1294-1296.

Bangera, R., Ødegård, J., Præbel, A.K., et al., 2011. Aquaculture. 317, 67-73.

Bangera, R., Odegård, J., Nielsen, H., et al., 2013. Journal of animal science.

Baranski, M., Kettunen Praebel, A., Sommer, A., et al., 2010. August 1-6, 2010 Leipzig, Germany. de.

Bush, W.S., Moore, J.H., 2012. PLoS computational biology. 8, e1002822.

Goddard, M., Hayes, B., 2007. Journal of Animal Breeding and Genetics. 124, 323-330.

Grisart, B., Coppieters, W., Farnir, F., et al., 2002. Genome research. 12, 222-231.

Hayes, B., Goddard, M., 2001. Genetics. 157, 1819-1829.

Houston, R.D., Haley, C.S., Hamilton, A., et al., 2008. Genetics. 178, 1109.

Moen, T., Baranski, M., Sonesson, A.K., et al., 2009. BMC Genomics. 10, 368.

Ødegård, J., Sommer, A.I., Præbel, A.K., 2010. Aquaculture. 300, 59-64.

Ødegård, J., Baranski, M., Gjerde, B., et al., 2011. Aquaculture Research. 42, 103-114.

Wedemeyer, G.A., 1996. Physiology of Fish in Intensive Culture Systems. Springer.