Aslam, M., Kettunen, A., Hansen, Ø., Lillehammer, M. (2026) Deciphering the genetic architecture of key economic traits in Atlantic cod breeding using a custom designed SNP Array, *Proceedings 13th World Congress on Genetics Applied to Livestock Production* 24039: https://doi.org/10.31274/wcgalp.24039 ©2026 Aslam, et al. CC BY-NC-ND

Deciphering the genetic architecture of key economic traits in Atlantic cod breeding using a custom designed SNP Array

M.L. Aslam*[1], A.H. Kettunen[12], Ø. Hansen[1], and M. Lillehammer[1]

[1]*Nofima, AS, Ås, Norway;* [2]*Nordic Genetic Resource Center, Ås, Norway;* **Corresponding author:* luqman.aslam@nofima.no

Introduction

The Atlantic cod (*Gadus morhua*) has long been one of Norway's most economically important fish species, supporting major capture fisheries and driving sustained efforts to develop a competitive aquaculture industry. In cod breeding programs, body weight is a key selection trait due to its direct impact on production efficiency and market value (Gjerde *et al*., 2004). However, early undesired sexual maturation remains a major biological constraint, reducing somatic growth, feed efficiency, and fillet quality (Trippel *et al*., 2014), while also raising ecological concerns due to the potential release of fertilized eggs from sea cages and associated risks to wild populations. Hence, achieving simultaneous genetic improvement in growth while delaying or controlling maturation is a one of the key challenges for sustainable Norwegian cod breeding.

Enhancing the competitiveness of cod aquaculture requires complementing traditional family-based selection with modern genomic approaches. Genomic selection, GS (Meuwissen *et al*., 2001) enables within family selection, increasing selection accuracy and intensity, and thereby accelerating genetic progress. This approach is particularly valuable for traits that are costly or logistically challenging to measure, such as disease resistance or early maturation, which often require large-scale experimental testing. The objective of this study was to develop a high-resolution medium density SNP array for Atlantic cod and to demonstrate its utility for estimation of genetic parameters, genome-wide association analyses (GWAS), and comparative prediction of breeding objective traits.

Materials & Methods

Sequence data analysis and selection of SNPs for development of array. Whole-genome resequencing data from 111 farmed (NFR Project No. 207680) and 30 wild Atlantic cod (Star et al., 2016) were used for SNP discovery and selection. Raw reads were quality-filtered with Trimmomatic v0.36 (Bolger et al., 2014) and retained if ≥ 36 bp after trimming. High-quality reads ($n \approx 141$ individuals) were aligned to the gadMor3.0 reference genome and processed using the GATK pipeline for variant discovery and annotation. SNPs selected for array development were filtered based on minor allele frequency (MAF), functional annotation (McLaren et al., 2016), inter-marker distance to reduce redundancy, and genome-wide coverage to ensure uniform end-to-end representation across chromosomes.

Resource population. The test population used for this study originated from the national cod breeding program of Norway and comprised of three year-classes, YCs (YC2016, YC2019, and YC2023). These sub-populations included 199 families from YC2016, 67 families from YC2019 and 111 families from YC2023. All year-classes were recorded for body weight (BW) while sexual maturation (SM; binary 0/1) trait was recorded on YC2019 and YC2023 through dissection of the fish and visual inspection of gonads. The recorded individuals were PIT-tagged, had pedigree information and were tissue sampled for DNA extraction and subsequent genotyping. Using the newly developed genotyping tool, 3,672 individuals were genotyped: 1,211 from YC2016, 1,264 from YC2019, and 1,197 from YC2023.

Estimation of genetic parameters. Genetic parameters were estimated within each year-class and combined data using following bivariate linear mixed model implemented in ASREML 4.2 (Gilmour *et al*., 2015),

$$\boldsymbol{y_1}\ \boldsymbol{y_2} = \mu + \boldsymbol{Xb} + \boldsymbol{Zu} + \boldsymbol{e}$$

where $y_1$and y_2 are the vector of phenotypes (SM and BW); $\boldsymbol{\mu}$ is the vector of trait means, $\mathbf{X}$ and $\mathbf{b}$ are the incidence matrix and vector of fixed effects of sex (YC included for combined year-class analysis), and $\mathbf{Z}$ relates observations to additive genetic effects $\mathbf{u}$. Additive genetic effects were assumed to follow $\mathbf{u} \sim \mathcal{N}(0, \mathbf{G} \otimes \mathbf{K})$, where $\mathbf{G}$ is the genetic (co)variance matrix between traits and $\mathbf{K}$ is either the pedigree-based relationship matrix or the genomic relationship matrix constructed using VanRaden's method 1. Residuals were assumed to be $\mathbf{e} \sim \mathcal{N}(0, \mathbf{R} \otimes \boldsymbol{I_n})$, where $\mathbf{R}$ is the residual (co)variance matrix and $\boldsymbol{I_n}$ is an identity matrix of order n

Genome-wide association analysis. Genome-wide association analysis for the sexual maturity trait was performed using a univariate linear mixed model implemented in GCTA with "*--mlma-loco*" option (Yang *et al*., 2011). The model structure was very similar to that described above, with the addition of individual SNP genotypes fitted as fixed marker effects to estimate allele-substitution effects.

Accuracy of prediction. Prediction accuracy was evaluated using a validation scheme in which 25% of sibling individuals were randomly selected as a validation set with masked phenotypes. Breeding values for these individuals were predicted using phenotypes from the remaining reference population together with either genomic or pedigree relationship matrices. Prediction accuracy was calculated mean across 40 replicates and expressed as the correlation between estimated breeding values and adjusted phenotypes, scaled by the square root of the corresponding heritability: $\boldsymbol{r_{corr}} = \frac{\boldsymbol{\rho(KEBV, y_{adj})}}{\sqrt{\boldsymbol{h^2}}}$, where $\boldsymbol{\rho}$ is the correlation coefficient, $\boldsymbol{KEBV}$ denotes genomic or pedigree-based breeding values, $\boldsymbol{y_{adj}}$the adjusted phenotypes, and $\boldsymbol{h^2}$the respective heritability estimate.

Results & Discussion

After initial filtering of raw biallelic SNPs based on base quality, read depth, and minor allele count, ~6.1 million SNPs were retained. Applying stricter thresholds (genotyping quality ≥ 30, MAF ≥ 0.15) reduced this set to ~900,000 SNPs. A final panel of ~22,000 SNPs was selected by considering chromosome-specific marker density, functional relevance, genome-wide coverage, and linkage disequilibrium. Following Illumina design requirements, 21,424 SNPs were incorporated into a GGP Illumina Infinium™ Atlantic cod array.

Genotype data obtained using new array was filtered applying MAF > 0.02 and ≥50% genotyping rate within each year class, 19,062 (YC2016), 18,673 (YC2019), and 18,687 (YC2023) SNPs were retained. Combining all datasets and filtering under the same criteria yielded 16,664 SNPs for downstream analyses. Sex information was recorded for all genotyped individuals, and a quick validation confirmed that the array detected the previously reported sex-associated quantitative trait loci (QTL), verifying its functionality and reliability (Star *et al*., 2016).

Phenotypic records were collected at age 2 for YC2019 and YC2023, and at 2.5 years for YC2016. Mean body weights (kg) were: 2.8 ± 0.60 (YC2016), 1.9 ± 0.52 (YC2019), and 1.7 ± 0.35 (YC2023). Heritability estimates for BW and SM obtained from pedigree-based and

genomic models for each year class and the combined dataset are summarized in Table 1. Body weight consistently showed moderate heritability, with joint analysis with estimates of 0.36 ± 0.04 (pedigree) and 0.37 ± 0.03 (genomic), aligning with reported BW heritability ranges of 0.3-0.5 in Atlantic cod (Gjerde *et al*., 2004; Kristjánsson and Arnason, 2015). Sexual maturity exhibited a high proportion of early-maturing fish in both year classes (85% in YC2019; 84% in YC2023). Across analyses, SM showed moderate heritability, though estimates varied with model type. Heritability estimates obtained through joint analysis were 0.24 ± 0.04 (pedigree LM) and 0.21 ± 0.03 (genomic LM), while threshold models produced higher estimates, 0.44 ± 0.08 (pedigree TM) and 0.51 ± 0.07 (genomic TM), consistent with the well-established tendency of linear models to underestimate heritability for binary traits on the liability scale (Gianola, 1982). As BW in sexually maturing and non-maturing fish showed only a moderate genetic correlation (0.56 ± 0.22), and the low frequency of non-maturing fish in the data is likely to introduce noise for genetic correlation estimates between SM and BW. With ~85% of fish scored for SM contributing BW records from maturing individuals, the estimated genetic correlations in this group ranged from 0.63 ± 0.08 (pedigree) to 0.51 ± 0.07 (genomic), indicating an unfavorable relationship. This pattern is consistent with previous findings (Kristjánsson and Arnason, 2015). However, these genetic correlation estimates should be re-evaluated using optimized timing of BW and SM recordings to prevent phenotypic effects of one trait from influencing the other.

Table 1. Year-class and full-dataset heritability estimates from genomic and pedigree models.

Dataset/Models	Pedigree Estimates (h^2)			Genomic Estimates (h^2)		
	LM		$TM_{sire-dam}$	***LM***		***TM***
	BW	**SM**	**SM**	**BW**	**SM**	**SM**
YC2016	0.34±0.07	NA±NA	NA±NA	0.45±0.05	NA±NA	NA±NA
YC2019	0.43±0.08	0.31±0.07	0.49±0.10	0.29±0.05	0.23±0.04	0.54±0.10
YC2023	0.39±0.07	0.15±0.05	0.33±0.10	0.33±0.06	0.20±0.05	0.51±0.12
All	0.36±0.04	0.24±0.04	0.44±0.08	0.37±0.03	0.21±0.03	0.51±0.07

LM = Estimates from a linear mixed model on the observed scale; TM = Liability-scale heritability estimates derived using a threshold sire–dam model or conversion; BW = Body weight; SM = Sexual maturity.

The Manhattan plot from the joint GWAS analysis of sexual maturity is shown in Figure 1. Given the low frequency of immature fish (~15.5%), combining the two year-class datasets and applying a customized statistical model helped enhance power to detect associated loci. The analysis identified a clear QTL on chromosome 20, with multiple SNPs exceeding the chromosome-wide significance threshold. This QTL remained consistent when males and females were analyzed separately, indicating a stable association across sexes. The lead SNP on chromosome 20 explained 5.5% of the additive genetic variance, estimated using the $2pq\alpha^2$method (Hill and Mackay, 2004). The limited proportion of contrasting phenotypes likely reduced the ability to detect loci of smaller effect; increasing sample size and achieving more balanced phenotype representation would likely refine QTL detection and effect estimation. Within the chromosome 20 QTL region, ***acvr1c*** and ***acvr1l*** emerge as strong candidate genes. As members of the ***acvr1*** family, these activin receptor genes are involved in TGF-β signaling and have established roles in puberty onset, gonadal differentiation, embryonic development, and growth-related processes (Valer et al., 2019; Shalev et al., 2025). Their presence in this region adds biological support to the involvement of this locus in regulating sexual maturation.

Prediction accuracies clearly demonstrated the advantage of using genomic over pedigree information. On average genomic models increased prediction accuracy by 30% for BW and 19% for SM over pedigree-based models. For BW, genomic and pedigree accuracies were

0.60 ± 0.07 and 0.46 ± 0.04; for SM, they were 0.61 ± 0.09 and 0.51 ± 0.05 respectively. These improvements indicate that implementing genomic information in selection programs should substantially enhance overall genetic gain and accelerate stock improvement for both traits.

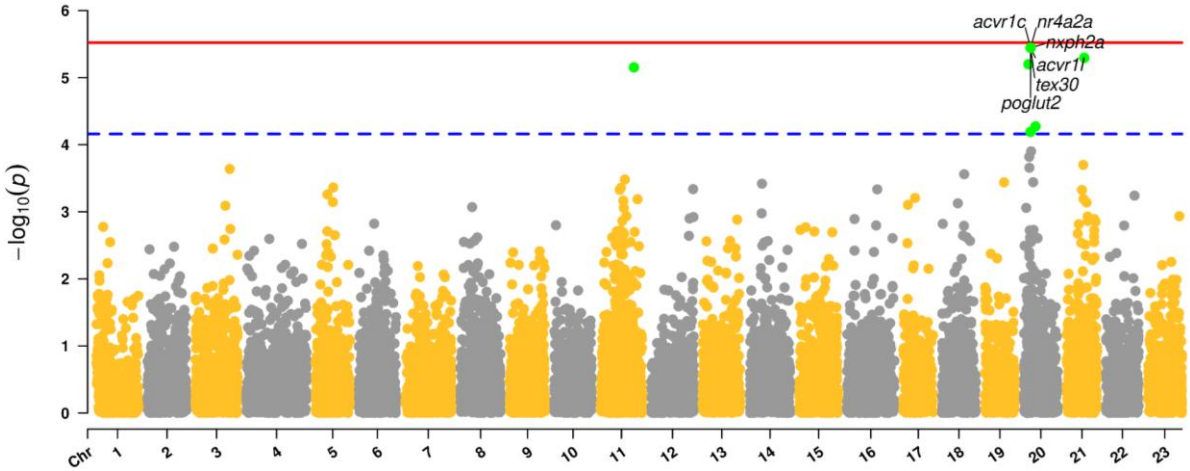

Figure 1: Manhattan plot presenting association of SNPs with the SM trait. The solid red line denotes the genome-wide Bonferroni threshold, and the dashed blue line indicates the chromosome-wide threshold. Genes within ±250 kb of the lead SNP on chromosome 20 are annotated.

Conclusions

This study demonstrates functionality of a ~21K SNP array for Atlantic cod, enabling robust genomic analyses of key economic traits. Both body weight and sexual maturity exhibited moderate heritability, confirming their suitability for genetic improvement. A stable QTL for sexual maturity was detected on chromosome 20, with biologically relevant candidate genes. Genomic prediction consistently outperformed pedigree-based approaches, indicating strong potential for accelerated genetic gain.

Acknowledgements

The results presented in this study form part of the STRAMA project (Project No. 342635), the MABIT program (Project No. AF0102), and the National Atlantic Cod Breeding Program.

References

Gianola, D. 1982. Journal of Animal Science 54(5):1079–1096. doi: 10.2527/jas1982.5451079x

Gilmour, A. R., B. J. Gogel, B. R. Cullis, *et al*. 2015. ASReml User Guide Release 4.2 Structural Specification. VSN International Ltd, Hemel Hempstead, UK.

Gjerde, B., B. F. Terjesen, Y. Barr, *et al*. 2004. Aquaculture 236(1):167–177. https://doi.org/10.1016/j.aquaculture.2004.03.004

Hill, W. G., and T. F. C. Mackay. 2004. Genetics 167(4):1529–1536. https://10.1093/genetics/167.4.1529

Kristjánsson, T., and T. Arnason. 2015. Aquaculture Research 46(9):2185–2193. https://doi.org/10.1111/are.12377

Meuwissen, T. H. E., B. J. Hayes, and M. E. Goddard. 2001. Genetics 157(4):1819.

Star, B., O. K. Tørresen, A. J. Nederbragt, *et al*. 2016. Scientific Reports 6(1):31235. doi: 10.1038/srep31235

Trippel, E. A., I. A. E. Butts, A. Babin, *et al*. 2014. Journal of Experimental Marine Biology and Ecology 451:35–43. https://doi.org/10.1016/j.jembe.2013.10.030

Yang, J., S. H. Lee, M. E. Goddard, *et al*. 2011. Am J Hum Genet 88(1):76–82. doi: 10.1016/j.ajhg.2010.11.011