



## Development of genomic markers associated to growth-related traits and sex determination in lumpfish (*Cyclopterus lumpus*)

Alejandro P. Gutierrez<sup>a</sup>, Sarah-Louise Counter Selly<sup>a</sup>, Samuel M. Pountney<sup>a,b</sup>, John B. Taggart<sup>a</sup>, Panagiotis Kokkinias<sup>a</sup>, Thomas Cavrois-Rogacki<sup>c</sup>, Eduardo Jimenez Fernandez<sup>c</sup>, Herve Migaud<sup>a</sup>, Ingrid Lein<sup>d</sup>, Andrew Davie<sup>a</sup>, Michaël Bekaert<sup>a,\*</sup>

<sup>a</sup> Institute of Aquaculture, Faculty of Natural Sciences, University of Stirling, Stirling FK9 4LA, Scotland, UK

<sup>b</sup> University of Victoria, Victoria, BC V8P 5C2, Canada

<sup>c</sup> Otter Ferry SeaFish Ltd., Otter Ferry, Tighnabruaich PA21 2DH, Scotland, UK

<sup>d</sup> Nofima AS, Sjølsengvegen 22, Sunndalsøra 6600, Norway

### ARTICLE INFO

#### Keywords:

Lumpfish  
QTL  
GWAS  
Aquaculture

### ABSTRACT

Cleaner fish species have gained great importance in the control of sea lice, among them, lumpfish (*Cyclopterus lumpus*) has become one of the most popular. Lumpfish life cycle has been closed, and hatchery reproduction is now possible, however, current production is reliant on wild caught broodstock to meet the increasing demand. Selective breeding practices are called to play an important role in the successful breeding of most aquaculture species, including lumpfish.

In this study we analysed a lumpfish population for the identification of genomic markers linked to production traits. Sequencing of RAD libraries allowed us to identify, 7193 informative markers within the sampled individuals. Genome wide association analysis for sex, weight, condition factor and standard length was performed. One single major QTL region was identified for sex, while nine QTL regions were detected for weight, and three QTL regions for standard length.

A total of 177 SNP markers of interest (from QTL regions) and 399 high  $F_{st}$  SNP markers were combined in a low-density panel, useful to obtain relevant genetic information from lumpfish populations. Moreover, a robust combined subset of 29 SNP markers (10 associated to sex, 14 to weight and 18 to standard length) provided over 90% accuracy in predicting the animal's phenotype by machine learning. Overall, our findings provide significant insights into the genetic control of important traits in lumpfish and deliver important genomic resources that will facilitate the establishment of selective breeding programmes in lumpfish.

### 1. Introduction

Sea lice infestation remains the most pressing issue affecting salmon aquaculture worldwide. Losses linked to sea lice were estimated at €700 million worldwide in 2015 and continues to increase [1]. These losses not only result from reduced production due to sea lice-associated mortalities, decreased fish growth, and reduced flesh quality, but also from the cost of treatment against sea lice. This often involves the use of parasiticide chemicals or mechanical treatments of limited effectiveness or carry other health risks [2].

To address this issue, biological control of sea lice infection in Atlantic salmon cages has become an important alternative to tackle the one of the most important diseases affecting salmon aquaculture. This

strategy has gained increasing popularity mainly due to its effectiveness and environmental safety [3]. In Norway, for example, about 0.7 million cleaner fish were deployed in salmon cages in 2006 which drastically increased to 43 million cleaner fish deployed in 2019, while showing a slight decrease in the last couple of years (Norwegian Directorate of Fisheries, 2022 [4]). Estimated figures indicate that approximately half of the cleaner fish used in the Atlantic salmon industry are lumpfish (*Cyclopterus lumpus*), and the remainder includes diverse species of wrasse [5,6] such as ballan wrasse (*Labrus bergylta*), corkwing wrasse (*Symphodus melops*) and Goldsinny wrasse (*Ctenolabrus rupestris*). However, current cleaner fish production relies heavily on wild-caught broodstock, and production from farmed broodstock remains minimal [5,7].

\* Corresponding author.

E-mail address: [michael.bekaert@stir.ac.uk](mailto:michael.bekaert@stir.ac.uk) (M. Bekaert).

<https://doi.org/10.1016/j.ygeno.2023.110721>

Received 3 May 2023; Received in revised form 19 September 2023; Accepted 25 September 2023

Available online 26 September 2023

0888-7543/© 2023 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

In recent years, hatchery production technologies for cleaner fish, particularly lumpfish and ballan wrasse, have been under intense development. Among these species, lumpfish hatchery production has proven to be more straightforward, with relatively high and stable survival rates [1] being achieved, making it a promising candidate for extensive use as a cleaner fish [8]. Lumpfish is a sub-Arctic species commonly found along the Icelandic, Norwegian, and British coastlines, as well as the East coast of North America [9]. Crucially, the lumpfish life cycle has been closed, and hatchery reproduction is now possible. However, there are only a limited number of hatcheries producing lumpfish in Europe, and breeding programmes are notably lacking [1]. Selective breeding is an effective strategy for improving the production of aquaculture species, by enhancing economically important traits [10]. Lumpfish hatchery production, which imitates natural conditions, takes only about seven months before the fish are ready for deployment. This is significantly shorter than the roughly 1.5 years required for ballan wrasse deployment, making the production cycle much more cost-effective [1]. Contrary to most aquaculture species, lumpfish grow much faster than would be preferred, leading to problems associated with delousing behaviour [11]. There is a marked decline in delousing activity in lumpfish upon reaching a large body size (over 300 g in 6-10 months), likely due to their slow movement, feeding off salmon pellet, and increasing aggressive (and territorial) behaviour, mainly triggered by the onset of sexual maturation [11]. Therefore, the establishment of breeding programmes that allow the production of stocks with a more favourable growth rate and other desirable traits would greatly benefit lumpfish production.

Selective breeding of commercially important traits has been applied in the improvement of numerous aquaculture species [12]. Growth-related traits in particular are common targets for selection in aquaculture breeding programmes due to their impact in production and profitability, and have shown great improvement after a few generations [13,14]. This success in growth improvement can be attributed the medium to high heritability values that most aquaculture species show for these type of traits, ranging from 0.3 to 0.6 [12]. Contrary to the prevailing progression towards the growth rate increase of livestock species, lumpfish selection aims to reduce the growth if the animal to increase its usage as cleaner fish.

Recent developments in genomic technologies have transformed selective breeding programmes for aquaculture species. This has mostly been driven by continuous advances in sequencing technologies that enable high-throughput discovery and screening of genetic markers, in particular single nucleotide polymorphisms (SNPs), which are highly abundant and widely distributed through the genome [12]. Screening of thousands of SNPs via genotyping-by-sequencing (GBS) techniques (such as ddRAD) or by SNP arrays has become common practice in devising and managing selective breeding schemes for many commercial aquaculture species, including emerging ones [12,15]. GBS techniques are based on a reduced representation of the genome achieved by the use of restriction enzymes that will cut the DNA at specific sites and will create a library of sequences dispersed across the genome. The main advantage of GBS tools is that can provide sequencing of numerous individuals at enough depth to identify SNPs confidently. While high coverage whole genome sequencing can provide much better representation of the genome, the cost associated to the sequencing is still highly expensive, considering the of hundreds of animals needed for GWAS or breeding programme applications [15]. These technologies have also helped in the development of reference genome sequences for numerous aquaculture species [16], including the recent release of a reference genome for lumpfish that serves as a valuable genomic resource for advancing production of this species [17]. All these genomic resources have provided a keystone for the development of selection schemes by the application of marker assisted selection (MAS), the adoption of accurate genomic relationship matrices (GRM) for breeding value calculations, and the advancement of genomic selection in most aquaculture species [12].

To support the establishment of sustainable selective breeding programmes for lumpfish, this study aimed to develop genetic resources for the species and identify genomic regions associated with growth traits and sex. These efforts will contribute to the production of stocks with favourable growth rates and other traits of interest, ultimately leading to the development of an effective and sustainable solution to the sea lice infestation issue in salmon aquaculture.

## 2. Materials and methods

### 2.1. Family creation

Wild broodstock were obtained from Skjerneset Fisk at Averøy, Norway. A total of 14C. *lumpus* independent full-sib families (1Q to 14Q) were created, within 3 h of each other on the same day in October 2018, and reared from fertilisation to final sampling at the NOFIMA Cleaner Fish Unit at Sunndalsora, Norway. All families were reared in discrete incubation units/tanks as outlined below, that were supplied by a common water source to ensure comparable environmental conditions. During egg incubation water temperature averaged  $7.4 \text{ }^{\circ}\text{C} \pm 0.8 \text{ }^{\circ}\text{C}$ , while during larval rearing water temperature averaged  $11.5 \text{ }^{\circ}\text{C} \pm 0.5 \text{ }^{\circ}\text{C}$ . Following hatching, temperatures were increased from the incubation temperature to the larval rearing temperature over a period of one week. From fertilisation to just prior to hatching, egg masses were held within individual perforated trays ( $20 \times 30 \times 10 \text{ cm}$ ) submerged within a communal trough. Just prior to hatching these trays were transferred to individual 250 L cylindroconical tanks. Hatching in all families initiated within three calendar days of each other, no later than 300 degree-days post fertilisation. Larvae were fed following routine commercial practice, first with live feed (*Artemia*), before weaning to a commercial formulated feed (Otohime, PTAqua, Norway). At 90 days post hatch, when fish reached an average of  $0.58 \text{ g}$  wet weight, the total number of families being reared was rationalised to four, rearing capacity was limited which required the reduction of families and therefore four families were selected based on survival performance to this point to ensure a suitable number of study animals to complete the study. Stock numbers were balanced to an average of 3250 juveniles which were selected at random within each family. At this stage the four selected families were transferred into 1500 L cylindroconical tanks for the remainder of the study. The final four families (3Q, 7bQ, 10Q & 11Q), were reared following normal commercial practice until 180 days post hatch, when the final phenotyping sampling was performed.

### 2.2. Phenotype measurement and sampling

For all four families the same sampling regime was followed. A total of 100 individuals were selected at random, culled by lethal anaesthesia and then for each individual total length ( $\pm 1 \text{ mm}$ ), standard length ( $\pm 1 \text{ mm}$ ), weight ( $\pm 0.01 \text{ g}$ ) and sex (where identifiable) were recorded. Thereafter a family specific upper and lower size threshold was calculated (bottom 10% of population curve “small” and top 10% of population curve “big”) as shown in Table 1. Then, a further 100 “small” and 100 “big” individuals within each family were sampled. For all fish body condition was measured using Fulton's condition factor ( $K = 100 \times \text{weight}/\text{length}^3$ ; [18]). The subsequent genomic analysis was based on a 125 individual (50 “big”, 50 “small” as well as 25 random selected individuals) from each family. Gender of juveniles within each family was balanced where possible. A total of 36 wild broodfish (including 14 parental pairs) were also samples but collecting fine clips.

### 2.3. DNA extraction

Fin clips for all parents and offspring were stored in 99% ethanol at  $4 \text{ }^{\circ}\text{C}$  until DNA extraction. Genomic DNA was extracted using a salt extraction method as described before [19]. Total nucleic acid content and quality (260 nm/230 nm and 260 nm/280 nm ratios) were

**Table 1**

Summary of population statistics. Mean weight (g) and length (mm) distributions as well as maximum and minimum sizes observed, along with threshold sizes (as defined by individual total length) which demarked population specific upper (90%) and lower (10%) size thresholds for selective sampling.

	Wet weight (g)	Total length (mm)	Largest Individual	Smallest Individual	90% & 10% threshold
Family	9.76 ±	57.59 ±	57.13 g	0.81 g	>71 mm
3Q	7.7	13.4	105 mm	30 mm	<41 mm
Family	5.44 ±	46.99 ±	51.25 g	0.47 g	> 66 mm
10Q	5.2	12.5	107 mm	25 mm	< 33 mm
Family	3.64 ±	39.84 ±	43.82 g	0.25 g	> 57 mm
11Q	5.7	12.5	97 mm	21 mm	< 29 mm
Family	2.79 ±	39.53 ±	20.14 g	0.36 g	> 50 mm
7bQ	0.8	8.7	77 mm	22 mm	< 31 mm

determined by spectrometry (Nanodrop; Thermo Scientific, Hemel Hempstead, UK) before measuring double-stranded DNA concentrations using a Qubit dsDNA Broad Range Assay Kit and Qubit Fluorometer (Invitrogen, Paisley, UK).

#### 2.4. Library preparation and sequencing

The ddRAD libraries were prepared using an adapted version of an existing protocol [19]. Briefly, DNA from each sample was digested at 37 °C, for 75 min with restriction enzymes *Pst*I and *Nla*III (New England Biolabs, UK), followed by heat-inactivation at 65 °C, for 25 min. The DNA samples were then individually barcoded through the ligation of specific P1 and P2 adapters, each containing a unique five or seven base nucleotide sequence. After addition of pre-mixed adaptors (*Pst*I:*Nla*III 1:16) and incubation of samples at 22 °C, for 10 min, T4 ligase (2000 ceU/μg DNA), rATP (100 mM) and CutSmart buffer (1×) were added and samples incubated for 90 min at 22 °C, followed by heat inactivation (65 °C, 20 min). Libraries were column purified (PCR MinElute, Qiagen, Manchester, UK), size selected by gel electrophoresis (550-650 bp) and amplified by PCR (15 cycles). Sequencing was performed by Novogene (UK) Co. Ltd. (Cambridge, UK) using an Illumina Novaseq 6000 platform (150-base paired-end reads).

#### 2.5. Marker assembly and genotyping

The sequence data from the 536 individuals (4 families for 125 offspring, and 36 broodfish; Supplementary Table S1) were pre-processed to discard low quality reads (i.e., with an average quality score <20). Sequences lacking the restriction site or having ambiguous barcodes were discarded during sample demultiplexing stage. Retained reads were then aligned against the genomic assembly of *C. lumpus* (NCBI Assembly accession GCA\_009769545.1) using bowtie2 v2.3.5.1 [20] and assembled using *gstacks* from Stack v2.60 [21] (using the *marukilow* model, and ignoring the read pairing).

All loci that were common to at least two individuals, with no further filtering, were exported from Stacks. As the analysis was based on a ddRAD sequencing method, many SNPs were sequenced in a very limited number of individual, creating a lot of missing genotypes. To avoid to introduces errors or reliable result in the downstream analysis, we didn't use an imputation step on such large matrix, where a substantial portion of the genotypes is missing. Using PLINK v1.9 [22], groups of variants that shared the same coordinates were identified, and only the first marker was retained (*-list-duplicate-vars suppress-first*), to avoid duplications or indistinctions. Moreover, SNPs with unknown position or located in partial chromosomes were excluded from the analysis. For each dataset (parents and offspring, corresponding to two separate sampling campaigns), SNPs and individuals were further filtered for quality control in a two steps process, again using PLINK. First, SNP inclusion was confined to those with minor allele frequency ≥

0.005 and  $p$ -value of  $\chi^2$  test for Hardy-Weinberg equilibrium  $\geq 10^{-6}$  Then SNPs and animals with a call rate  $\geq 0.9$  were selected. Quality control was performed on the datasets (parents and offspring) independently. Filtered scores were then combined in one dataset, keeping only shared SNPs.

#### 2.6. Multidimensional scaling analysis

R v4.2.0 [23] was used to carry out Multidimensional Scaling Analysis on the dataset using the package Bioconductor/SNPrelate v1.30.1 [24] to calculate the Identity-By-State (IBS) proportion for each sample.

#### 2.7. Identification of trait associated markers

Using the recorded phenotypic data (total length, standard length, weight, condition factor and sex) association analyses were performed within the package R/SNPassoc v2.0-11 [25] for R v4.2.0 [23], using the “log-additive” model (except for sex, where “co-dominant” model was used). We used a  $p$ -value threshold of 0.001 and a corrected  $p$ -value for multiple tests of 0.001/number of tests. The model used for the analysis was based on Interval Mapping. The algorithm used considers the phenotype to follow a mixture of Bernoulli distributions and uses a form of the EM algorithm for obtaining maximum likelihood estimates [26]. Two-way and multiple quantitative trait locus (QTL) models were also run with this package. R/qrt2 v0.28 [27] was used to identify the confidence interval of the QTLs on each chromosomes. After calculating the kinship matrix with *calc\_kinship*, we performed a univariate QTL mapping using the *scan1* function for each trait.

#### 2.8. Genetic relationship matrix

A genetic relationship matrix (GRM) was used to estimate the genetic relationships between individuals from the SNPs. The assigned genetic relationship value between two individuals ( $i$  and  $k$ ) represents the off-diagonal values of the GRM and can be estimated by the following equation:

$$A_{jk} = \frac{1}{N} \sum_{i=1}^N \frac{(x_{ij} - 2p_i)(x_{ik} - 2p_i)}{2p_i(1 - p_i)}$$

Where  $x_i$  is the number of copies of the reference allele (0,1,2) for the  $i^{\text{th}}$  SNP and for individual  $j$  or  $k$  respectively,  $N$  is the number of SNPs in common (for both individuals  $j$  and  $k$ ) and  $p_i$  is the reference allele frequency of the  $i^{\text{th}}$  SNP. For each SNP the estimate is summed ( $\Sigma$ ) and weighed equally ( $1/N$ ).

The diagonal values of the GRM represent one plus the inbreeding coefficient and can be calculated with the following equation:

$$A_j = 1 + \frac{1}{N} \sum_{i=1}^N \frac{x_{ji}^2 - (1 + 2p_i)x_{ji} + 2p_i^2}{2p_i(1 - p_i)}$$

#### 2.9. Estimation of the variance explained by genome-wide SNPs by REML

A restricted maximum likelihood estimation (REML) analysis was used to partition the variance of phenotypic trait across a group of observed values (fixed effects) and the genomic relatedness (random effect based on the GRM). The analysis was performed in the GCTA v1.94.1 [28] using the following equation:

$$Y = X\beta + Z\gamma + \varepsilon$$

where  $Y$  is the phenotype,  $X$  is the incidence matrix for fixed effects,  $\beta$  is the fixed variable effect size,  $Z$  is the GRM,  $\gamma$  is the vector of random effects from the GRM, and  $\varepsilon$  is the residual random effect (representing environmental, non-genetic effects).

## 2.10. Estimating components of heritability

With the fixed and random (GRM) effects specified, the model is fitted using REML to generate an estimate of variance explained by the GRM adjusted for all the fixed effects. GCTA calculates by default a likelihood ratio test (LRT) in order to examine the significance of the random effects for the GRM on the fit of the model ( $p$ -value). SNP-based heritability estimates were obtained by the proportion of genotypic to phenotypic variation obtained by the GRM in the analysis.

## 2.11. QTL strength model

The effect of all SNP markers for each QTL was analysed using WEKA v3.8 [29], which contains a variety of machine-learning algorithms, including “REPTree” [29], a fast decision tree learner that builds a decision/regression tree using information gain/variance and reduced-error pruning with backfitting. “REPTree” considers all the markers, then derives for each individual a phenotype prediction (lengths, weight and sex) based on its genotypes for the markers considered. To train the model, we used minimum total weight of the instances in a leaf of 3 (*minNum*) and the amount of data used for pruning was 4 (*numFold*). The most predictive SNP markers for each QTL were selected and used to produce a reduced SNP panel with the same prediction power compared to the full set of markers. Permutatively, individuals were removed one-by-one from the training set, with the algorithm subsequently assigning their predicted phenotypic values.

## 2.12. Low density SNP panel

To develop an extensive SNP panel able to capture all genomic regions of interest, as well as maximising the estimation of diversity, all SNP markers associated with the phenotypes of interest (lengths, weight, and sex) were selected as well as markers with the highest  $F_{st}$  values (elevated genetic differentiation).  $F_{st}$  were calculated using the function *gl.fst.pop* from dartR v2.7.2 [30] and based on all available allele frequencies for each families and broodfish. After several tests run by LGC Genomics (Teddington, UK), SNP markers that presented successful amplification by SeqSNP were used for the final panel.

## 2.13. Panel validation

The usefulness of the SNP panel was validated to confirm the association of the selected markers to the analysed traits. For this purpose, additional members of the four families previously used in the genome-wide association study (GWAS) analyses, as well as parents, were genotyped, selecting the rest of the “big” and “small” individuals from each family (398 samples [~100 samples from each family], as well as a further 59 broodfish and 20 duplicated samples for control; Supplementary Table S2). In total, tissue samples from 477 fish were shipped and genotyped by LGC Genomics (Teddington, UK).

## 3. Results

### 3.1. Library sequencing

High throughput sequencing of 536 individuals produced 3,260,920,744 paired-end reads in total. After the removal of low-quality and incomplete reads, 78.9% of the total raw reads were retained (2,571,378,028 PE-reads; Supplementary Table S1). *C. lumpus* genome was used to map the reads and generate ddRAD-tags. A total of 3,048,066 unique loci were detected, with 477,421 shared by at least two samples.

### 3.2. SNP identification and quality control

From the 477,421 SNPs initially identified between the two groups

(36 parents and 500 offspring), the filtering process left 35 parents with 19,227 SNPs passing the threshold, and 499 offspring with 8186 SNPs, as shown in Table 2. A total of 7193 common informative markers were identified (covering the remaining 534 individuals) and used in subsequent analyses (Supplementary Data S1 and Supplementary Table S3).

### 3.3. Sample structure

A Multidimensional Scaling Analysis (IBS) was utilized to capture the complex structure of the samples and separate the individuals into clusters based on their genetic distance [31]. This process grouped individuals of same origins together (families), while positioning prior family assignment errors or poor-quality samples as outliers (Fig. 1). Five distinct clusters were separated using the first two components (67.3% of cumulative variance). Families and parental/wild generation were clearly clustered. There was one exception; individual 11Q-212 did not behave as expected and did not cluster with any of the families, most likely due to wrong assignment during sampling or handling issues during the transfer into family tanks.

### 3.4. Association analysis

Making use of the 7193 QC filtered and informative SNP markers, R/ SNPpass and R/qtl2 were used to conduct a QTL/GWAS analysis for both sex and morphometric ratios measurements. Genome wide association was detected for both sex and morphometric measurements after Bonferroni test correction (Supplementary Table S3). One single major QTL (57 SNPs) was identified for sex determination (Fig. 2A), whereas a total of nine QTL regions (120 SNPs) were detected for weight (Fig. 2B), and three QTL regions (23 SNPs) for standard length (Fig. 2C). On the other hand, no significant association was detected when using Fulton's condition factor as a trait (Fig. 2D). All SNP markers associated with standard length were also significantly associated with Weight (Fig. 3 and Table 3). The SNP based heritability estimates for all traits showed high values, with length showing a  $h^2$  of 0.57 ( $\pm 0.04$ ), condition factor 0.49 ( $\pm 0.06$ ) and 0.67 ( $\pm 0.04$ ) for weight; Sex showed a heritability of 1.

### 3.5. Prediction and validation

The combined prediction power of these 177 unique SNP markers (from the sex and weight QTL) was tested by building machine-learning algorithms and using an instance-based  $k$ -nearest neighbours' algorithm [32] based on the additive effect of the genotypes at each locus considered. Phenotype prediction power from using these 177 SNPs provided a 99% accuracy for the selection of a desired phenotype (Table 4). The SNP markers defining the QTLs for Weight and Standard length were further investigated to provide a small subset of marker fit for a quick SNP assay. This approach produced a robust combined subset of 29 SNP markers (10 associated to sex, 14 to weight and 18 to standard length, with weight and standard-length markers overlapping; Supplementary Table S4). When applied to all individuals, the combined prediction power remains over 90% (Table 4).

### 3.6. Low density SNP panel

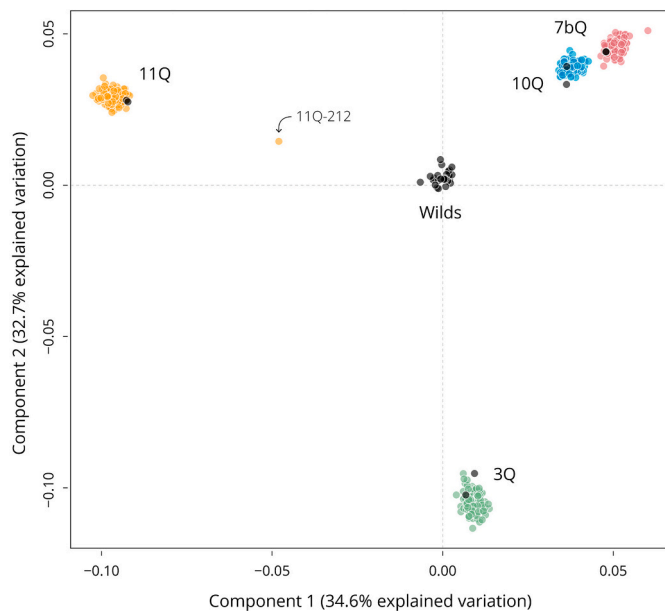
A total of 177 SNP markers of interest (from QTL regions) and 399 top  $F_{st}$  SNP markers were combined in a low-density panel to test its usefulness in quantifying and maintaining genetic diversity within the tested population, along with potential use for selection purposes in the future. The inclusion the high  $F_{st}$  SNP also allow for detecting local adaptation, and merits inclusion in future genome scans and methodologies. This final panel of 576 markers successfully delivering informative genotypes was selected (Supplementary Data S2) and evaluated on the previously mentioned 477 samples, showing its usefulness to provide genotypes that can be used for selection purposes and family



**Table 2**

SNP markers filtering steps. \*First step: SNP CR (Call rate) = 0.1, HWE (Hardy-Weinberg equilibrium) =  $10^{-6}$ . \*\*Second step: Animal CR (Call rate) = 0.1, MAF (Minor allele frequency) = 0.005.

Population	Initial		Excluded				Remaining	
	Animal	SNP	HWE*	SNP CR*	MAF**	Animal CR**	Animal	SNP
Parents	36	453,181	14	417,273	16,667	1	35	19,227
Offspring	500	453,181	2521	418,009	24,465	1	499	8161



**Fig. 1.** Multidimensional Scaling Analysis results of the full dataset. The first and second components explain 34.6%, and 32.7% of the variation found. Based on 7193 SNP markers. The black dots clustering with each family are the two brookstock/wild parents. Families 7bQ and 10Q are more clearly separated in the component 3 (not show in this figure).

structure (Fig. 4), although lacking enough resolution to determine population structure.

## 4. Discussion

### 4.1. Importance of generating genomic resources for emerging species

The development of comprehensive genomic resources for important and emerging aquaculture species is crucial for gaining deeper insights into their biological characteristics, optimising breeding programmes, and enhancing selective breeding for desirable traits [12,15]. Genomic tools can be particularly advantageous for emerging species such as lumpfish, as it can expedite and improve the accuracy of the selection process for important traits, in addition to establishing breeding programmes.

In this study, genomic markers were developed for a lumpfish stock, with 7193 informative SNPs being identified following a thorough QC filtering process. This is a significant achievement and represents a valuable starting point for future genomic research on the species. Genomic resources, such as DNA markers, have become an essential component of successful aquaculture production. As a result, many relevant aquaculture species have been targeted for SNP marker development via genotyping-by-sequencing (GBS) or whole genome sequencing [15,16].

The development of SNP markers has been extensively researched in Atlantic salmon over the last two decades, starting with the screening of a few hundred SNPs, to reach the development and application of numerous SNP arrays containing up to 900K markers [33–37]. Sea

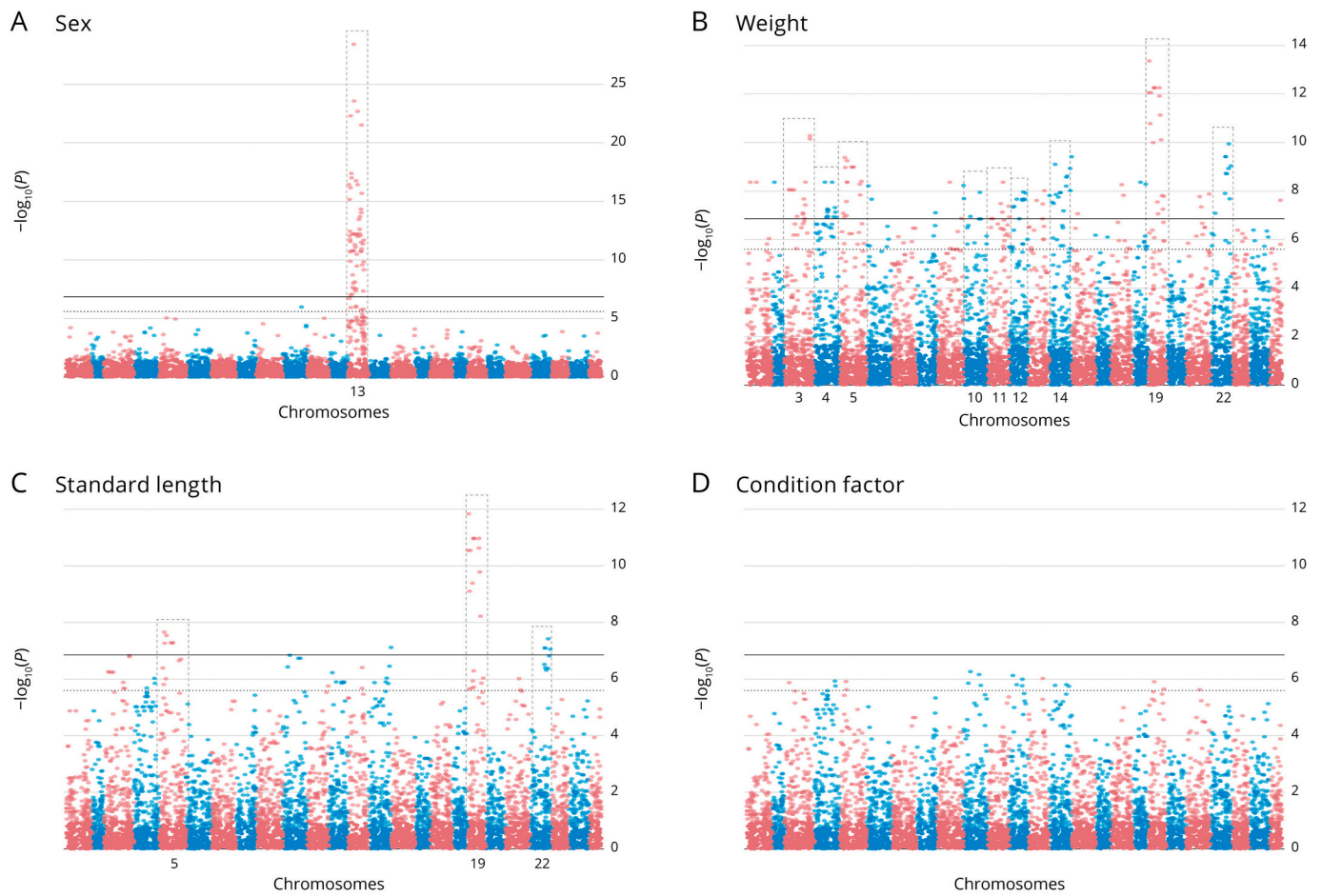
bream and sea bass are additional examples of species with successful genomic marker development, transitioning from GBS-based SNP identification to the development of medium-density SNP arrays [38–40]. Low-density SNP panels have emerged as a cost-effective alternative, providing access to genotyping resources at a fraction of the cost compared to high-density panels, especially when dealing with hundreds or thousands of animals. These panels have been successfully described for numerous aquaculture species and have demonstrated their application in genomic selection through accurate genomic relationship matrices (GRM) or imputation into high-density panels [41].

This progress exemplifies how commercial interest and production needs can positively stimulate research advancement in aquaculture species, which could serve as a model for lumpfish research given the current demand for cleaner fish.

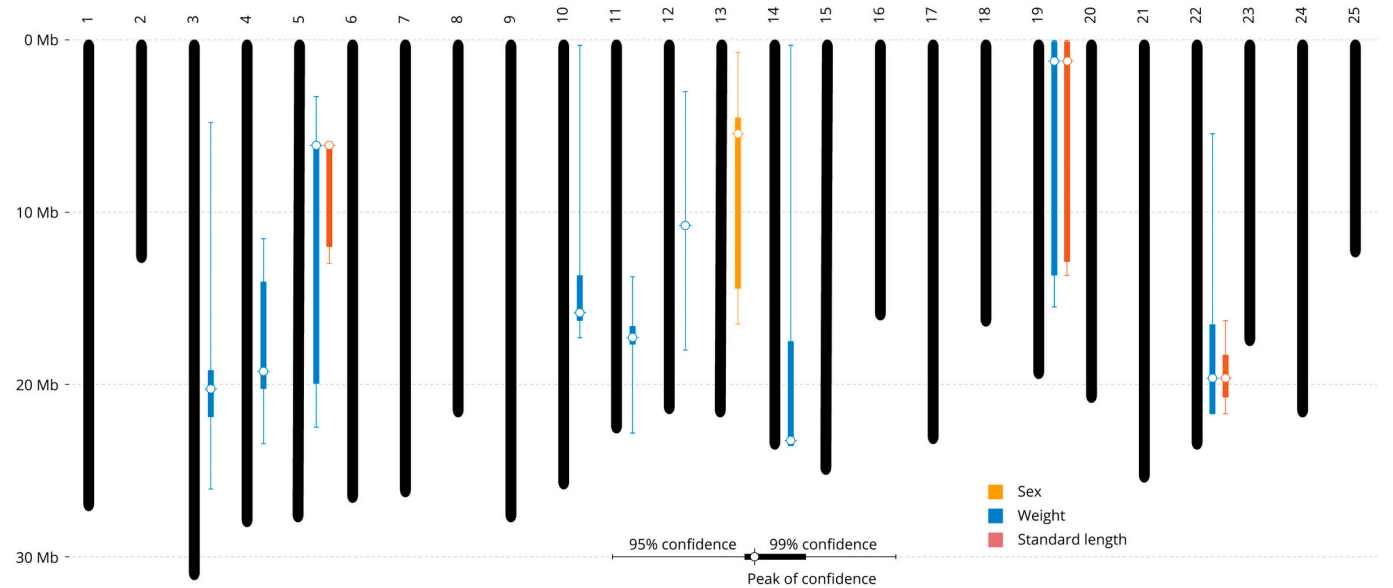
### 4.2. GWAS on growth and sex in other species

The increased accessibility to sequencing technologies has made GWAS (and QTL) analyses involving thousands of markers a norm for studying aquaculture and livestock species. This has enabled the identification of significant associations between genomic markers and particular phenotypes, such as growth, sex, disease resistance, and colour, among others, through GWAS and QTL analyses, which is a fundamental step towards the selective improvement of stocks. Aquaculture species have been extensively researched for the identification of QTL regions and markers associated with important traits [42]. The present study aimed to identify genomic regions associated with analysed traits, and was successful in this regard (Table 3). The analysis of sex showed the strongest association, with a single major QTL located on chromosome 13 (Fig. 2A and Fig. 3) being identified. This result is in line with what was recently reported for another lumpfish stock, where chromosome 13 was also identified as the sex chromosome, and the *Amh* gene was suggested as the sex determining gene [17]. The *Amh* gene, short for “Anti-Müllerian hormone” gene, plays a significant role in sex determination and differentiation in various vertebrate species, including several species of fish, either as a primary determinant, a secondary influencer, or as part of a broader network of genes and molecular pathways. Although the QTL peak position does not exactly match the position of the *Amh* gene in the lumpfish genome. This could be due to many reasons, first the previous study utilized a 70K SNP array for lumpfish, while our analyses were based on 7193 SNPs, and therefore, even though our results agree with the chromosome location, the lower marker density did not provide enough resolution to identify the specific location of the candidate gene. Nevertheless, a set of 10 SNPs located within this region accurately predicted sex in all samples, giving evidence that the sex determining locus is shared between populations.

Growth rate is a significant trait for improvement in newly domesticated species, and it has been extensively studied in most aquaculture species [42]. Analysis of weight and length in lumpfish showed polygenic involvement, identifying significant associations across many chromosomes, and showing evidence of overlap in QTL regions in chromosomes 5, 19, 14, and 22 (Fig. 2B–C, Table 3, and Fig. 3). In contrast, the analysis of condition factor (K) did not identify significant associations, most likely due to the round morphology of lumpfish, which makes this index uninformative for this species [43]. The polygenic nature of growth traits is not surprising, as most aquaculture species show this pattern. Contrary to most reported aquaculture goals,



**Fig. 2.** Markers associated with phenotypes. (A) Manhattan plot of the association for phenotypic sex. (B) Manhattan plot of the association with the fish weight. (C) Manhattan plot of the association with the fish Standard length. (D) Manhattan plot of the association with the fish condition factor. The  $-\log_{10}(p\text{-value})$  values for association of directly genotyped SNPs are plotted as a function of position of the physical map. Each chromosome has been represented with a different colour.



**Fig. 3.** QTL map. Chromosomal locations of highlighted genomic regions for QTLs in this study, including Standard length, fish weight, phenotypic sex and condition factor. The peak locations are located with white circles.

**Table 3**

QTL detected in this study and the genomic regions harbouring them. The peaks and confidence intervals (CI) can be visualised in Fig. 2, while the QTLs are reported in Fig. 3. \*Small peaks not reported in Fig. 3.

Traits	QTL	Chr.	Confidence intervals (bp)	Peak (LOD; bp)	Markers	Mean of LOD
Sex	SexQTL	13	888,322 - 16,609,738	29.5; 5,952,081	57	12.8
Weight	QTL1*	1	3,584,426 - 9,963,809	8.3; 3,584,426	2	8.3
	QTL2*	2	1,931,736 - 1,931,736	8.3; 1,931,736	1	8.3
	QTL2	3	4,805,596 - 26,754,368	10.2; 26,754,368	12	8.1
	QTL4	4	11,522,283 - 23,529,190	8.3; 19,116,224	13	7.3
	QTL5	5	3,362,532 - 22,429,695	9.3; 4,965,412	15	8.3
	QTL6*	6	1,992,512 - 5,395,107	8.2; 1,992,512	2	7.9
	QTL7*	8	20,552,086 - 20,552,086	7.1; 20,552,086	1	7.1
	QTL8*	9	13,608,747 - 25,957,741	8.3; 13,608,747	2	7.6
	QTL9	10	249,201 - 17,391,982	8.3; 249,201	4	7.7
	QTL10	11	13,911,073 - 22,899,415	8.3; 16,341,550	5	7.4
	QTL11	12	4,710,808 - 18,097,392	7.9; 18,097,392	9	7.6
	QTL12*	13	8,434,979 - 16,572,657	8.0; 16,572,657	3	7.7
	QTL13	14	260,826 - 23,505,589	9.4; 23,505,589	15	8.1
	QTL14*	15	6,396,440 - 6,396,440	7.2; 6,396,440	1	7.2
	QTL15*	17	11,464,082 - 13,337,808	8.3; 11,464,082	2	8.1
	QTL16*	18	8,262,460 - 13,642,938	8.3; 8,262,460	2	8.3
	QTL17	19	470,829 - 16,571,612	13.2; 1,188,533	17	10.4
	QTL18*	21	14,317,822 - 24,347,228	7.9; 24,347,228	3	7.6
	QTL19	22	5,460,636 - 21,864,352	9.4; 16,287,751	10	8.6
	Std length	QTL20*	25	11,069,039 - 11,069,039	7.5; 11,069,039	1
QTL1		5	4,965,412 - 13,898,772	7.6; 4,965,412	6	7.3
QTL2*		14	23,505,589 - 23,505,589	7.1; 23,505,589	1	7.1
QTL3		19	470,829 - 13,758,119	11.7; 1,188,533	12	10.2
QTL4		22	16,287,751 - 21,864,352	7.4; 19,726,257	4	7.1

**Table 4**

Details of the Phenotypic Variation Explained and prediction accuracy for the full SNP dataset and reduced subset. For each trait tested, the subset of SNPs is reported between brackets. The marker subsets overlap. Sex is a binary character, where correct prediction is provided, Weight and Std length are continuous variables where Precision (Correlation) is specified. Subset list is provided in Supplementary Table S4 (29 unique SNP markers).

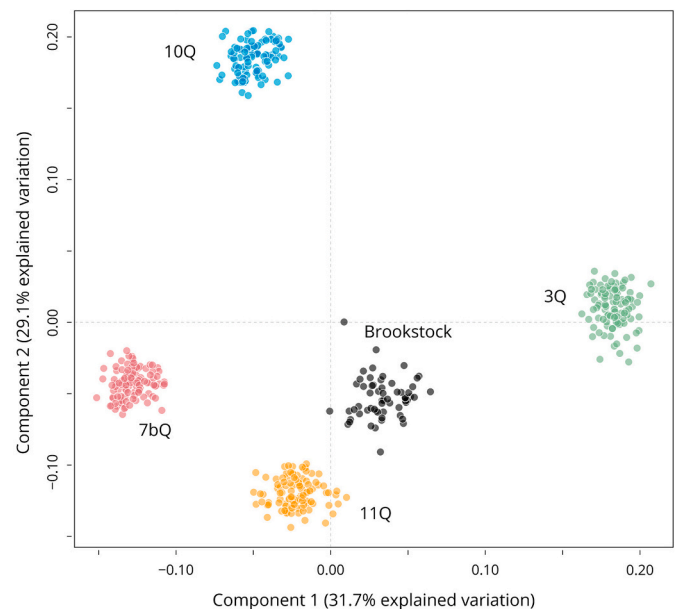
	Markers	Correct prediction Precision	Mean absolute error
Sex (All)	177	99.7%	0.008
Sex (subset)	10	95.5%	0.0742
Weight (All)	177	99.9%	10.257 g
Weight (subset)	14	92.5%	67.468 g
Std length (All)	177	99.8%	1.433 mm
Std length (subset)	18	91.4%	16.410 mm

where growth rate QTLs have been largely exploited to increase growth rate, selection for slower growth rate and longer deployment time may be feasible for lumpfish. Grazing efficacy has been negatively correlated with the size of lumpfish and linked to parental/family effects, suggesting that the genetic component can play a significant role in improving growth and grazing [44]. Accordingly, obtained heritability estimates (0.57, 0.49, 0.67 and 1 for length, condition factor, weight and Sex respectively) are in accordance to values estimated for growth traits in other aquaculture species [12], therefore, improvement in growth rate (for better grazing efficiency) can be anticipated, considering that the genetic component plays an important role in the expression of these traits.

The markers identified in this study show promise for the selection of slow-growing fish using a low number of markers (Table 4), and therefore, they have the potential to improve the grazing efficiency of selected stocks.

#### 4.3. Usefulness of findings. MAS and genomic selection applications

The QTL markers identified in this study have great potential to significantly enhance the analysed traits, particularly growth, which has shown average genetic gains of over 10% per generation in some



**Fig. 4.** Multidimensional Scaling Analysis results of the validation panel (477 samples) dataset. The first and second components explain 31.7%, and 29.1% of the variation found. Based on 576 SNP markers.

aquaculture species [45]. Selection to obtain the opposite outcome (slower growth) should be possible at similar rates, particularly with the introduction of genomic resources into the selection process. Our analysis of 177 markers associated with the traits revealed 99% accuracy in predicting the animal's phenotype, and a selection of only 29 SNPs achieved similar accuracy, thus opening the possibility of using low-density SNP panels, such as the one described in this study, to provide practical genomic resources at a lower cost without sacrificing selection power.

The results of this study demonstrate that a panel of 576 markers can determine family structure and accurately predict slow growth phenotypes, and sex-associated markers can accurately distinguish the sex of

individuals, which is particularly beneficial for selecting broodstock at early stages (Table 4 and Fig. 4). Furthermore, these genomic resources can be used to determine relatedness, population structure, genetic variation, and inform genomic selection [41]. The inclusion of high SNP  $F_{st}$  marker will also allow for detecting local adaptation in future genome scans and methodologies [46]. While further analyses are necessary to explore the SNP panel's ability to differentiate the geographical origin of lumpfish populations and test the application of genomic selection for improving selection schemes, the findings provide significant insight into the genetic control of important traits in lumpfish.

Overall, the developed genomic resources offer great potential for facilitating the development of breeding programmes for lumpfish and selection based on genomic information. Our study sheds light on the genetic factors influencing growth and sex in lumpfish and highlights the potential of low-density SNP panels as a cost-effective and powerful tool for genomic selection in aquaculture.

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.ygeno.2023.110721>.

### Ethics statement

Ethical approval for the study was granted in accordance with the Norwegian state Mattilsynet programme and in accordance with the UK Animals (Scientific Procedures) Act 1986 Amended Regulations (SI 2012/3039), as well as approved by the Animal Welfare and Ethical Review Body of the University of Stirling (AWERB (18 19) 032).

### Funding

This research was funded by EU funded AquaExcel 2020, Trans-national access grant (AE100022), the Biotechnology and Biological Sciences Research Council and Natural Environment Research Council [grant number BB/S004416/1] and the Sustainable Aquaculture Innovation Centre.

### CRedit authorship contribution statement

**Alejandro P. Gutierrez:** Conceptualization, Formal analysis, Investigation, Methodology, Project administration, Resources, Supervision, Writing – original draft, Writing – review & editing. **Sarah-Louise Counter Selly:** Conceptualization, Funding acquisition, Methodology, Resources, Writing – review & editing. **Samuel M. Pountney:** Methodology, Writing – review & editing. **John B. Taggart:** Methodology, Writing – review & editing. **Panagiotis Kokkinias:** Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Validation, Writing – review & editing. **Thomas Cavois-Rogacki:** Funding acquisition, Resources, Writing – review & editing. **Eduardo Jimenez Fernandez:** Writing – review & editing. **Herve Migaud:** Conceptualization, Funding acquisition, Writing – review & editing. **Ingrid Lein:** Conceptualization, Funding acquisition, Writing – review & editing. **Andrew Davie:** Conceptualization, Funding acquisition, Methodology, Resources, Writing – review & editing. **Michaël Bekaert:** Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Supervision, Validation, Writing – original draft, Writing – review & editing.

### Declaration of Competing Interest

T.C.R. and E.J.F. are employed by Otter Ferry SeaFish Ltd., a funding partner of the research project. The other authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

### Data availability

All the raw reads sequence data obtained in this study were deposited at the European Bioinformatics Institute (EBI) Sequence Read Archive (SRA) project ID PRJEB38408.

### References

- [1] A.J. Brooker, A. Papadopoulou, C. Gutierrez, S. Rey, A. Davie, H. Migaud, Sustainable production and use of cleaner fish for the biological control of sea lice: recent advances and current challenges, *Vet. Rec.* 183 (2018), <https://doi.org/10.1136/vr.104966>, 383–383.
- [2] M.J. Costello, The global economic cost of sea lice to the salmonid farming industry, *J. Fish Dis.* 32 (2009) 115–118, <https://doi.org/10.1111/j.1365-2761.2008.01011.x>.
- [3] M.J. Costello, Ecology of sea lice parasitic on farmed and wild fish, *Trends Parasitol.* 22 (2006) 475–483, <https://doi.org/10.1016/j.pt.2006.08.006>.
- [4] Norwegian Directorate of Fisheries, Sale of Farmed Cleaner Fish 2012–2021. <https://www.fiskeridir.no/English/Aquaculture/Statistics/Cleanerfish-Lumpfish-and-Wrasse>, 2022 (accessed April 1, 2023).
- [5] M. Bolton-Warberg, An overview of cleaner fish use in Ireland, *J. Fish Dis.* 41 (2018) 935–939, <https://doi.org/10.1111/jfd.12731>.
- [6] K. Overton, L.T. Barrett, F. Oppedal, T.S. Kristiansen, T. Dempster, Sea lice removal by cleaner fish in salmon aquaculture: a review of the evidence base, *Aquacult. Environ. Interact.* 12 (2020) 31–44, <https://doi.org/10.3354/aei00345>.
- [7] A. Powell, J.W. Treasurer, C.L. Pooley, A.J. Keay, R. Lloyd, A.K. Imsland, C. Garcia de Leaniz, Use of lumpfish for sea-lice control in salmon farming: challenges and opportunities, *Rev. Aquac.* 10 (2018) 683–702, <https://doi.org/10.1111/raq.12194>.
- [8] A.K.D. Imsland, A. Hanssen, A.V. Nytrø, P. Reynolds, T.M. Jonassen, T.A. Hangstad, T.A. Elvegård, T.C. Urskog, B. Mikalsen, It works! Lumpfish can significantly lower sea lice infestation in large-scale salmon farming, *Biol. Open.* 7 (2018), <https://doi.org/10.1242/bio.036301> bio036301.
- [9] J. Davenport, Synopsis of Biological Data on the Lumpfish, *Cyclopterus lumpus* (Linnaeus, 1758), Food and Agriculture Organization of the United Nations, Rome, Italy, 1985. <https://www.fao.org/3/ap950e/ap950e.pdf>.
- [10] T. Regan, T.P. Bean, T. Ellis, A. Davie, S. Carboni, H. Migaud, R.D. Houston, Genetic improvement technologies to support the sustainable growth of UK aquaculture, *Rev. Aquac.* 13 (2021) 1958–1985, <https://doi.org/10.1111/raq.12553>.
- [11] A.K. Imsland, P. Reynolds, A.V. Nytrø, G. Eliassen, T.A. Hangstad, Ó.D. B. Jónsdóttir, P.-A. Emaus, T.A. Elvegård, S.C.A. Lemmens, R. Rydland, T. M. Jonassen, Effects of lumpfish size on foraging behaviour and co-existence with sea lice infected Atlantic salmon in sea cages, *Aquaculture.* 465 (2016) 19–27, <https://doi.org/10.1016/j.aquaculture.2016.08.015>.
- [12] R.D. Houston, T.P. Bean, D.J. Macqueen, M.K. Gundappa, Y.H. Jin, T.L. Jenkins, S. L.C. Selly, S.A.M. Martin, J.R. Stevens, E.M. Santos, A. Davie, D. Robledo, Harnessing genomics to fast-track genetic improvement in aquaculture, *Nat. Rev. Genet.* 21 (2020) 389–409, <https://doi.org/10.1038/s41576-020-0227-y>.
- [13] T. Gjedrem, N. Robinson, Advances by selective breeding for aquatic species: a review, *AS.* 05 (2014) 1152–1158, <https://doi.org/10.4236/as.2014.512125>.
- [14] K. Janssen, H. Chavanne, P. Berentsen, H. Komen, Impact of selective breeding on European aquaculture, *Aquaculture.* 472 (2017) 8–16, <https://doi.org/10.1016/j.aquaculture.2016.03.012>.
- [15] D. Robledo, C. Palaiokostas, L. Bargelloni, P. Martínez, R. Houston, Applications of genotyping by sequencing in aquaculture breeding and genetics, *Rev. Aquac.* 10 (2018) 670–682, <https://doi.org/10.1111/raq.12193>.
- [16] G.H. Yue, L. Wang, Current status of genome sequencing and its applications in aquaculture, *Aquaculture.* 468 (2017) 337–347, <https://doi.org/10.1016/j.aquaculture.2016.10.036>.
- [17] M.K. Holborn, A.L. Einfeldt, T. Kess, S.J. Duffy, A.M. Messmer, B.L. Langille, M. K. Brachmann, J. Gauthier, P. Bentzen, T.M. Knutsen, M. Kent, D. Boyce, I. R. Bradbury, Reference genome of lumpfish *Cyclopterus lumpus* Linnaeus provides evidence of male heterogametic sex determination through the AMH pathway, *Mol. Ecol. Resour.* 22 (2022) 1427–1439, <https://doi.org/10.1111/1755-0998.13565>.
- [18] T.W. Fulton, Rate of growth of seas fishes, in: *20th Ann. Rep. Fish. Board Scotland, 1902*, pp. 326–446.
- [19] J.K. Brown, J.B. Taggart, M. Bekaert, S. Wehner, C. Palaiokostas, A.N. Setiawan, J. E. Symonds, D.J. Penman, Mapping the sex determination locus in the hpku (*Polyprion oxygenios*) using ddRAD sequencing, *BMC Genomics* 17 (2016) 448, <https://doi.org/10.1186/s12864-016-2773-4>.
- [20] B. Langmead, S.L. Salzberg, Fast gapped-read alignment with bowtie 2, *Nat. Methods* 9 (2012) 357–359, <https://doi.org/10.1038/nmeth.1923>.
- [21] J.M. Catchen, A. Amores, P. Hohenlohe, W. Cresko, J.H. Postlethwait, Stacks: Building and genotyping loci *de novo* from short-read sequences, *G3* 1 (2011) 171–182, <https://doi.org/10.1534/g3.111.000240>.
- [22] S. Purcell, B. Neale, K. Todd-Brown, L. Thomas, M.A.R. Ferreira, D. Bender, J. Maller, P. Sklar, P.I.W. de Bakker, M.J. Daly, P.C. Sham, PLINK: a tool set for whole-genome association and population-based linkage analyses, *Am. J. Hum. Genet.* 81 (2007) 559–575, <https://doi.org/10.1086/519795>.
- [23] R Core Team, R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing, 2022.



- [24] X. Zheng, D. Levine, J. Shen, S.M. Gogarten, C. Laurie, B.S. Weir, A high-performance computing toolset for relatedness and principal component analysis of SNP data, *Bioinformatics*. 28 (2012) 3326–3328, <https://doi.org/10.1093/bioinformatics/bts606>.
- [25] J.R. González, L. Armengol, X. Solé, E. Guinó, J.M. Mercader, X. Estivill, V. Moreno, SNPAssoc: an R package to perform whole genome association studies, *Bioinformatics*. 23 (2007) 644–645, <https://doi.org/10.1093/bioinformatics/btm025>.
- [26] K.W. Broman, S. Sen, *A Guide to QTL Mapping with R/Qtl*, Springer, New York, NY, 2009, <https://doi.org/10.1007/978-0-387-92125-9>.
- [27] K.W. Broman, D.M. Gatti, P. Simecek, N.A. Furlotte, P. Prins, S. Sen, B.S. Yandell, G.A. Churchill, R/qtl2: software for mapping quantitative trait loci with high-dimensional data and multiparent populations, *Genetics*. 211 (2018) 495–502, <https://doi.org/10.1534/genetics.118.301595>.
- [28] J. Yang, S.H. Lee, M.E. Goddard, P.M. Visscher, GCTA: a tool for genome-wide complex trait analysis, *Am. J. Hum. Genet.* 88 (2011) 76–82, <https://doi.org/10.1016/j.ajhg.2010.11.011>.
- [29] I.H. Witten, E. Frank, M.A. Hall, C.J. Pal, *The WEKA workbench*, in: *Data Mining: Practical Machine Learning Tools and Techniques*, Fourth, Morgan Kaufmann, Cambridge, MA, 2017, pp. 553–571, <https://doi.org/10.1016/B978-0-12-804291-5.00024-6>.
- [30] B. Gruber, P.J. Unmack, O.F. Berry, A. Georges, dartR: An R package to facilitate analysis of SNP data generated from reduced representation genome sequencing, *Mol. Ecol. Resour.* 18 (2018) 691–699, <https://doi.org/10.1111/1755-0998.12745>.
- [31] I.T. Jolliffe, J. Cadima, Principal component analysis: a review and recent developments, *Philos. Trans. A Math. Phys. Eng. Sci.* 374 (2016) 20150202, <https://doi.org/10.1098/rsta.2015.0202>.
- [32] D.W. Aha, D. Kibler, M.K. Albert, Instance-based learning algorithms, *Mach. Learn.* 6 (1991) 37–66, <https://doi.org/10.1007/BF00153759>.
- [33] T. Moen, B. Hayes, M. Baranski, P.R. Berg, S. Kjøglum, B.F. Koop, W.S. Davidson, S. W. Omholt, S. Lien, A linkage map of the Atlantic salmon (*Salmo salar*) based on EST-derived SNP markers, *BMC Genomics* 9 (2008) 223, <https://doi.org/10.1186/1471-2164-9-223>.
- [34] A.P. Gutierrez, K.P. Lubieniecki, E.A. Davidson, S. Lien, M.P. Kent, S. Fukui, R. E. Withler, B. Swift, W.S. Davidson, Genetic mapping of quantitative trait loci (QTL) for body-weight in Atlantic salmon (*Salmo salar*) using a 6.5 K SNP array, *Aquaculture* 358–359 (2012) 61–70, <https://doi.org/10.1016/j.aquaculture.2012.06.017>.
- [35] R.D. Houston, J.B. Taggart, T. Cézard, M. Bekaert, N.R. Lowe, A. Downing, R. Talbot, S.C. Bishop, A.L. Archibald, J.E. Bron, D.J. Penman, A. Davassi, F. Brew, A.E. Tinch, K. Gharbi, A. Hamilton, Development and validation of a high density SNP genotyping array for Atlantic salmon (*Salmo salar*), *BMC Genomics* 15 (2014) 90, <https://doi.org/10.1186/1471-2164-15-90>.
- [36] J.M. Yáñez, S. Naswa, M.E. López, L. Bassini, K. Correa, J. Gilbey, L. Bernatchez, A. Norris, R. Neira, J.P. Lhorente, P.S. Schnable, S. Newman, A. Mileham, N. Deeb, A.D. Genova, A. Maass, Genomewide single nucleotide polymorphism discovery in Atlantic salmon (*Salmo salar*): validation in wild and farmed American and European populations, *Mol. Ecol. Resour.* 16 (2016) 1002–1011, <https://doi.org/10.1111/1755-0998.12503>.
- [37] M. Sinclair-Waters, J. Ødegård, S.A. Korsvoll, T. Moen, S. Lien, C.R. Primmer, N. J. Barson, Beyond large-effect loci: large-scale GWAS reveals a mixed large-effect and polygenic architecture for age at maturity of Atlantic salmon, *Genet. Sel. Evol.* 52 (2020) 9, <https://doi.org/10.1186/s12711-020-0529-8>.
- [38] C. Palaiokostas, M. Bekaert, J.B. Taggart, K. Gharbi, B.J. McAndrew, B. Chatain, D. J. Penman, M. Vandeputte, A new SNP-based vision of the genetics of sex determination in European sea bass (*Dicentrarchus labrax*), *Genet. Sel. Evol.* 47 (2015) 68, <https://doi.org/10.1186/s12711-015-0148-y>.
- [39] C. Palaiokostas, S. Ferraresso, R. Franch, R.D. Houston, L. Bargelloni, Genomic prediction of resistance to Pasteurellosis in gilthead sea bream (*Sparus aurata*) using 2b-RAD sequencing, *G3* 6 (2016) 3693–3700, <https://doi.org/10.1534/g3.116.035220>.
- [40] C. Peñaloza, T. Manousaki, R. Franch, A. Tsakogiannis, A.K. Sonesson, M.L. Aslam, F. Allal, L. Bargelloni, R.D. Houston, C.S. Tsigonopoulos, Development and testing of a combined species SNP array for the European seabass (*Dicentrarchus labrax*) and gilthead seabream (*Sparus aurata*), *Genomics*. 113 (2021) 2096–2107, <https://doi.org/10.1016/j.ygeno.2021.04.038>.
- [41] C. Kriaridou, S. Tsairidou, R.D. Houston, D. Robledo, Genomic prediction using low density marker panels in aquaculture: performance across species, traits, and genotyping platforms, *Front. Genet.* 11 (2020) 124, <https://doi.org/10.3389/fgene.2020.00124>.
- [42] J.M. Yáñez, A. Barriá, M.E. López, T. Moen, B.F. García, G.M. Yoshida, P. Xu, Genome-wide association and genomic selection in aquaculture, *Rev. Aquac.* 15 (2023) 645–675, <https://doi.org/10.1111/raq.12750>.
- [43] C. García de Leaniz, C. Gutierrez Rabadan, S.I. Barrento, R. Stringwell, P.N. Howes, B.A. Whittaker, J.F. Minett, R.G. Smith, C.L. Pooley, B.J. Overland, L. Biddiscombe, R. Lloyd, S. Consuegra, J.K. Maddocks, P.T.J. Deacon, B.T. Jennings, S. Rey Planellas, A. Deakin, A.I. Moore, Daniel Phillips, G. Bardera, M.F. Castanheira, M. Scolamacchia, N. Clarke, O. Parker, J. Avizienius, M. Johnstone, M. Pavlidis, Addressing the welfare needs of farmed lumpfish: Knowledge gaps, challenges and solutions, *Rev. Aquac.* 14 (2022) 139–155, <https://doi.org/10.1111/raq.12589>.
- [44] A.K.D. Inmsland, P. Reynolds, T.A. Hangstad, L. Kapari, S.N. Maduna, S.B. Hagen, Ó. D.B. Jónsdóttir, F. Spetland, K.S. Lindberg, Quantification of grazing efficacy, growth and health score of different lumpfish (*Cyclopterus lumpus* L.) families: Possible size and gender effects, *Aquaculture* 530 (2021) 735925, <https://doi.org/10.1016/j.aquaculture.2020.735925>.
- [45] T. Gjedrem, N. Robinson, M. Rye, The importance of selective breeding in aquaculture to meet future demands for animal protein: a review, *Aquaculture*. 350–353 (2012) 117–129, <https://doi.org/10.1016/j.aquaculture.2012.04.008>.
- [46] T. Da Silva Ribeiro, J.A. Galván, J.E. Pool, Maximum SNP  $F_{ST}$  outperforms full-window statistics for detecting soft sweeps in local adaptation, *Genome Biol. Evolut.* 14 (2022), <https://doi.org/10.1093/gbe/evac143> evac143.