Contents lists available at ScienceDirect

Aquaculture Reports

journal homepage: www.elsevier.com/locate/aqrep

Genomic prediction model optimization for growth traits of olive flounder (*Paralichthys olivaceus*)

W.K.M. Omeka^{a,b}, D.S. Liyanage^{a,b}, Sukkyoung Lee^{a,b}, H.M.V. Udayantha^a, Gaeun Kim^a, Subothini Ganeshalingam^a, Taehyug Jeong^{a,b}, David B. Jones^c, Cecile Massault^c, Dean R. Jerry^{c,d,*}, Jehee Lee^{a,b,**}

^a Department of Marine Life Sciences & Fish Vaccine Research Center, Jeju National University, Jeju Self-Governing Province 63243, Republic of Korea

^b Marine Life Research Institute, Jeju National University, Jeju Self-Governing Province 63333, Republic of Korea

^c ARC Hub for Supercharging Tropical Aquaculture through Genetic Solutions, James Cook University, Townsville, QLD 4811, Australia

^d Tropical Futures Institute, James Cook University, Singapore

ARTICLE INFO

Keywords: Genomic prediction Genomic relatedness Olive flounder Predictive ability Single nucleotide polymorphisms

ABSTRACT

Genomic prediction (GP) has emerged an effective tool for addressing the many shortcomings of traditional selective breeding, thereby enhancing the selection process. In this study, we optimized GP methods using 5-fold cross-validation to estimate genome-estimated breeding values for the weight traits of olive flounder (Paralichthys olivaceus). To accomplish our goal, we determined the parentage of the target broodstock and the ability of 11 prediction models to predict the weight traits of 1.8-year-old olive flounders, which were genotyped using a 70 K single nucleotide polymorphism (SNP) array. Moreover, our optimization efforts toward the predictive ability of genomic best linear unbiased prediction (GBLUP), Bayesian B (BB), and random forest (RF) methods encompassed changes in various aspects such as fixed effects, SNP quantity, population size, and phenotypic data collected at different fish ages. Additionally, we assessed the predictive ability for the total length and body depth of fish using GBLUP, BB, and RF. Among the 11 prediction methods used in this study, the BB (0.675), Elastic Net (0.679), and RF (0.698) methods exhibited the highest predictive abilities, whereas the GBLUP (0.637) method demonstrated the lowest. Incorporating information regarding fish sex as a fixed effect substantially improved the predictive ability of GBLUP and BB. For mean models, utilizing 3000-5000 random SNP markers resulted in a higher predictive ability, similar to that obtained using 50,000 SNPs. Increasing the population size reduced the standard deviation of the predictive ability. Notably, phenotypic records from 1.8year-old fish exhibited a significantly higher predictive ability than those from the other age groups. Furthermore, GBLUP, BB, and RF provided higher predictive abilities for length (0.655-0.852) and body depth (0.665–0.861). These findings may significantly shape future olive flounder genomic selection programs and offer valuable insights into GP in aquaculture.

1. Introduction

Olive flounder (*Paralichthys olivaceus*) is an important carnivorous flatfish species in South Korea, where it accounts for 70% of the global supply. Flounder is commercially valuable for various reasons, including its good flavor, disease resistance, and early economic turnover. Although interdisciplinary activities have been undertaken to boost the output of flounder aquaculture, considerable mortality has occurred owing to the widespread dissemination of new diseases and

environmental variables (Jung et al., 2020; Sohn et al., 2019). Therefore, selecting the best-performing fish for the aquaculture industry has been identified as a necessity and has been promoted through research.

Genomic prediction (GP) estimates the impact of all loci using genome-wide markers and forecasts the genome-estimated breeding value (gEBV) of the progeny. Therefore, dispersing markers over the genome such that at least one marker is in linkage disequilibrium (LD) with each quantitative trait locus (QTL) is necessary (Meuwissen et al., 2001). Incidentally, single nucleotide polymorphism (SNP) markers

https://doi.org/10.1016/j.aqrep.2024.102132

Received 15 April 2024; Accepted 6 May 2024

Available online 11 May 2024

2352-5134/© 2024 The Author(s). Published by Elsevier B.V. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).







^{*} Corresponding author at: ARC Hub for Supercharging Tropical Aquaculture through Genetic Solutions, James Cook University, Townsville, QLD 4811, Australia

^{**} Correspondence to: Marine Molecular Genetics Lab, Jeju National University, 102 Jejudaehakno, Jeju 63243, Republic of Korea

E-mail addresses: dean.jerry@jcu.edu.au (D.R. Jerry), jehee@jejunu.ac.kr (J. Lee).

have been successfully used to address these constraints.

Implementing various prediction approaches, such as best linear unbiased prediction (BLUP), Bayesian, and machine learning (ML) methods, allows for the prediction of animal phenotypes and gEBVs at very early stages, reduces production costs variously, and considerably improves selection decisions. The cross-validation method has been widely used to train models for genomic selection (Schrauf et al., 2021). In cross-validation, datasets are divided into reference and validation sets, the phenotype of the validation dataset is masked, and prediction is achieved by fine-tuning the model using the reference dataset. Model optimization can be realized by adjusting parameters such as marker density, population size, and prediction method.

Although GP has been used in many aquaculture species at the research scale, salmonid aquaculture has commercially employed it as a promising technique for selection (Chaivichoo et al., 2023; D'Agaro et al., 2021; Lu et al., 2023; Song et al., 2023, 2022; Verbyla et al., 2022). Conventional breeding strategies to improve growth parameters have been comprehensively studied in olive flounder (Kim et al., 2011; Li et al., 2019, 2018). In addition, Microsatellite (MS) marker- and pedigree-based selection methods have been evaluated experimentally (Kang et al., 2006). Recently, genome-wide association studies of olive flounder and other aquaculture species have been conducted for different traits, such as growth, viral hemorrhagic septicemia virus resistance, and thermal tolerance, using high-density SNP markers (Liyanage et al., 2022; Omeka et al., 2022; Udayantha et al., 2023; Wang et al., 2023). As growth is a highly polygenic trait, GP is a better approach for selecting individuals than marker-assisted or selective breeding. In addition, the moderate heritability of olive flounder weight can be favorable for genomic selection (Omeka et al., 2022). Therefore, we optimized the GP models of olive flounder weight as an approach to prediction-based genomic selection. To achieve this goal, we performed a parentage analysis for the selected fish; compared different prediction methods, including population size, SNP number, and the inclusion of fixed effects, in different scenarios; and identified the optimal parameters for prediction.

2. Methodology

2.1. Study population

This study used the same farmed population described previously (Omeka et al., 2022). Briefly, 37 maternal and 66 paternal candidates (oP) were obtained from different regions of Korea and Japan. In 2019, the study population (oF0_KW) was produced via strip spawning at the Ocean and Fisheries Research Institute (Jeju, South Korea). Fish were grown out at a local farm (Jeju, South Korea) in a single tank, wherein temperature was regulated by continuously flowing seawater (18-22 °C). Fish were fed moist pellets, and at approximately 1 year of age, they were tagged with Trovan chips (PIT; Trovan Ltd., Douglas, Isle of Man). At different age points (June 2020, 1 year; September 2020, 1.5 years; January 2021, 1.8 years; September 2021, 2.3 years), fish were randomly selected, and phenotypic measurements, including body weight, total length, and body depth, were taken. The sex of the fish was recorded during the gonadal maturation stage (January 2021 to September 2021). At the age of 1.8 years, the fin was clipped and stored at -80 °C until genomic DNA (gDNA) extraction. All the experimental steps were performed under the Flounder Genomic Selection Project and reviewed and approved by the Animal Care and Use Committee of Jeju National University (Approval Number: 2021-0033).

2.2. Genotyping and quality control

gDNA extraction, SNP panel design, genotyping, SNP quality filtering, and sample isolation were performed as previously described (Omeka et al., 2022). A total of 103 parent and 1045 oF0_KW offspring fin samples (50 mg) were used to extract the gDNA using a QIAamp 96

DNA QIAcube HT[™] kit and QIAcube HT automated DNA extractor (Qiagen, Hilden, Germany). Thereafter, the gDNA was diluted to 50 ng/µL and sent for genotyping at the Ramaciotti Center of Genomics (Sydney, NSW, Australia) using a 70 K Affymetrix Axiom® myDesign[™] SNP array (Affymetrix, Santa Clara, CA, USA). After quality control (QC) using Axiom® Analysis Suite (version 4.0; Affymetrix) and PLINK software (https://www.cog-genomics.org/plink/) (Chang et al., 2015), 103 oP and 1009 oF0_KW samples were successfully retained.

2.3. Parentage assignment and genomic relatedness matrix

Parentage assignment was performed in R software (R Foundation for Statistical Computing, Vienna, Austria) using 4000 SNP markers. The SNP data of both offspring (oF0_KW) and parents (oP) obtained from the Axiom Analysis Suite were converted to binary files using PLINK software. The genotype data were quality controlled at a >90% genotype rate (geno) and minor allele frequency (MAF) > 5%. Customized R codes were used to identify the parents of each individual using the opposite homozygote count method. The results were confirmed using Cervus v3.0. The genomic relatedness matrix was generated using the genomic relationship matrix (GRM) function in the GASTON package of R (http s://rdocumentation.org/packages/gaston/versions/1.4.9) and visualized using a heatmap.

2.4. Genomic prediction and cross-validation

The genotype data obtained from the ped and map files were converted to binary format using PLINK 1.9 software. Next, the data were quality-filtered based on genotyping rate (>90%) and MAF (>5%). Genotype and phenotype were matched for each individual. Missing genotypes were imputed by the MNI method, where missing data were replaced by the mean allele frequency of the SNP marker. These data were used in different models, methods, and scenarios to evaluate gEBV. Here, 5-fold cross-validation with three replications was performed, and the mean predictive ability was calculated as the Pearson correlation between the actual phenotype and gEBV.

2.5. Prediction methods

To identify the best prediction method for the weight of the oF0_KW fish, we used 11 different prediction methods, including genomic BLUP (GBLUP), extended GBLUP (EGBLUP), Bayesian A (BA), Bayesian B (BB), Bayesian C (BC), Bayesian Lasso (BL), Bayesian ridge regression (BRR), ridge regression (RR), elastic net (EN), reproductive kernel Hilbert space (RKHS), and random forest regression (RF). Different R packages were used for each method, as described below. The phenotypic records of 1009 fish at 1.8 years were used in this prediction. A basic mean model (without including the fixed effect) with 5-fold cross-validation and three replications was run in this step.

2.5.1. GBLUP

GBLUP is a widely used method for GP to estimate breeding values using genomic data (VanRaden, 2008). This involves solving mixed-model equations using a GRM. The key equation is as follows:

$$Y = Zg + X\beta + e$$

where "*Y*" is the vector of phenotypes, "*X*" is the design matrix for fixed effect " β ," "*Z*" is the incidence matrix for random effects "*g*" (breeding values), and "*e*" represents the residual error. We assumed that *g* and *e* follow a normal distribution as $g \sim N$ (0, $K\sigma_g^2$) and $e \sim N$ (0, $I\sigma_e^2$), respectively. "*K*" is the genomic relatedness matrix, and "*T*" is an identity matrix. The rrBLUP v 4.6.2 package (https://www.rdocumentation.or g/packages/rrBLUP/versions/4.6.2) with the kin.blup function was used for GBLUP prediction.

2.5.2. EGBLUP

The EGBLUP model includes additive and epistatic genetic effects (Jiang and Reif, 2015). The key equation is similar to that of GBLUP but with an additional term for incorporating external information:

$$Y = \mu + Zx1 + Wx2 + e$$

where " μ " is the overall mean of the trait and "e" is the residual error. "Z" and "W" are index matrixes of "X1" and "X2" respectively. "X1" is an ndimensional vector of additive genotypic values, and "X2" is an ndimensional vector of additive \times additive epistatic genotypic values. For EGBLUP prediction, the BWGS v 0.2.1 library (https://www.rdocumen tation.org/packages/BWGS/versions/0.2.1) was used for EGBLUP prediction.

2.5.3. BA, BB, BC, BL and BRR

BA, BB, BC, BL, and BRR were prepared as described previously (Lin et al., 2020; Meuwissen et al., 2001). The key equation for the Bayesian models is as follows:

$$Y = \mu \mathbf{1}_n + X\beta + \sum_{i=1}^p z_i g_i + e$$

where "*Y*," "*X*," " β ," and "e" are the same as in the equation for GBLUP. " μ " is the overall mean, and "1 *n*" is a vector consisting of ones; "*p*" signifies the total number of genotypes for an individual. " Z_i " denotes the vector of genotypes at the ith SNP. " g_i " stands for the vector of random effects, which represents the additive genetic effect for the ith SNP. BA employs a scaled *t*-distribution as its prior distribution, whereas BB employs a mixture of a Gaussian distribution and a point mass at zero in its prior distribution. BC utilizes a prior distribution consisting of a mixture of a scaled *t*-distribution and a point mass at zero. In contrast, the prior distributions, respectively. All Bayesian predictions were performed using the BGLR v 1.1.0 library (https://www.rdocumentation. org/packages/BGLR/versions/1.1.0) implemented in R.

2.5.4. Ridge regression

RR is a linear regression method that adds an L2 regularization term to a least-squares objective function (Ogutu et al., 2012). The key equation for the RR estimates of the β coefficients is as follows:

$$\widehat{\beta}_{ridge} = arg \ min_{\beta} \quad \left\{ ||y - X\beta|| \frac{2}{2} + \lambda ||\beta|| \frac{2}{2} \right\}$$

where "y" is the vector of dependent variables (phenotypes), "X" is the design matrix for independent variables, and " λ " is the regularization parameter. For RR prediction, the glmnet v 4.1.7 library (https://www.rdocumentation.org/packages/glmnet/versions/4.1–7) was used.

2.5.5. Elastic net

EN is a hybrid regression method that combines L1 and L2 regularization terms to achieve both sparsity and shrinkage of coefficients (Zou and Hastie, 2005). The key equation for the EN estimate of the β coefficients is as follows:

$$\widehat{\beta}_{EN} = arg \min_{\beta} \left\{ \left| \left| \mathbf{y} - \mathbf{X} \beta \right| \right| \frac{2}{2} + \lambda_1 \left| \left| \beta \right| \right|_1 + \lambda_2 \left| \left| \beta \right| \right| \frac{2}{2} \right\}$$

where "y" is the vector of observed phenotypes, "X" is the design matrix for independent variables, and " λ_1 " and " λ_2 " are regularization parameters controlling the strength of L1 and L2 regularization, respectively. For EN prediction, the glmnet v 4.1.7 library (https://www.rdocume ntation.org/packages/glmnet/versions/4.1–7) was used.

2.5.6. RKHS

RKHS is a mathematical framework used in kernel methods, including support vector machines and kernel RR (Gianola and Van Kaam, 2008). This involves mapping data points to a high-dimensional

space using a kernel function that allows for nonlinear modeling. The key equation for GP using RKHS is as follows:

$$f(\mathbf{y}) = \sum_{i=1}^{n} \alpha_i k(\mathbf{x}_{i,i})$$

where "f(y)" represents the predicted phenotype for an individual with genotype x, "n" is the number of training individuals in the data set. " α_i " is the coefficient associated with each individual, " x_i " represents the genotypes of the ith individual, " $k(x_i, x)$ " is the kernel function that measures the similarity or relationship between the genotypes of the ith training individual and the genotype x of the individual for whom prediction is required. For RKHS prediction, the glmnet v 4.1.7 library (https://www.rdocumentation.org/packages/glmnet/versions/4.1–7) was used.

2.5.7. Random forest regression

RF is an ensemble learning method that builds multiple decision trees and averages their predictions to make the final predictions (Breiman, 2001). In regression, the key equation involves averaging the predictions from the individual decision trees as follows:

$$\widehat{y} = \frac{1}{N_{trees}} \sum_{i=1}^{N_{trees}} tree_i(x)$$

where " \hat{y} " is the predicted value for a given input "x," "tree_i" is the prediction from the "ith" decision tree, and "N_{trees}" is the total trees. For RF, randomForest v 3.3–4 (https://www.rdocumentation.org/pac kages/randomForest/versions/3.3–4) was used.

2.6. Fixed effects model

Of the 11 prediction methods, GBLUP and the best-performing Bayesian model (BB) were used to include different fixed effects, such as family and sex, which optimized the model. Parental information was obtained as described in Section 2.4. The oFO_KW fish (1009 fish) phenotype records for January 2021 (1.8 years old) were used for 5-fold cross-validation.

2.7. Marker density

GBLUP and the BB and ML (RF) methods were used to identify the effect of marker number on predictive ability. In addition, both fixedeffects (GBLUP_fixed and BB_fixed) and mean models (GBLUP_mean, BB_mean, and RF) were used. Markers were divided into multiples of 100—up to 1000 SNP markers—and used for cross-validation. For 1000–10,000 SNPs, the markers were randomly divided into multiples of 1000. Subsequently, markers from 10,000–50,000 were divided into multiples of 10,000. SNP files were generated using the random sampling reduct.marker.size ((RMR) function in the BWGS v0.2.1 library (https://www.rdocumentation.org/packages/BWGS/versions/0.2.1). The total population was used, and 5-fold cross-validation was performed with three replications.

2.8. Population size and number of folds

The population (1009 fish) was randomly split into multiples of 100 (100, 200, 300, 400, 500, 600, 700, 800, 900, and 1000) by PLINK using a fam file, and each dataset was separately subjected to 5-fold cross-validation. The number of SNPs used for cross-validation was 54,730, and each cross-validation was repeated thrice. The predictive ability was estimated for GBLUP_mean, BB_mean, RF, GBLUP_fix, and BB_fix.

k-fold cross-validation (k = 1-10) was conducted to determine the effect of the number of folds on predictive ability. The total population and all QC-filtered SNPs were used to identify the fold of cross-validation, and each experiment was repeated thrice.

2.9. Age

Fish with phenotypic records were selected for all four measurement times (June 2020, 1.2 years old; September 2020, 1.5 years old; January 2021, 1.8 years old; September 2021, 2.3 years old). All QC-passed SNPs were used, and 5-fold cross-validation with three replications was performed for the GBLUP_mean, BB_mean, RF, GBLUP_fix, and BB_fix prediction formulas. Heritability of weight, length, and width at each age of fish was calculated as per the protocol outlined by Omeka et al. (2022).

2.10. Prediction of length and body depth of the fish

Prediction of total length and body depth was performed for the selected GBLUP_mean, BB_mean, RF, GBLUP_fix, and BB_fix methods. The corresponding phenotypic records of 1009 fish for January 2021 and the total number of SNPs were used with 5-fold cross-validation with three replications.

2.11. Mean squared error of prediction and statistical analysis

The mean squared error of prediction (MSEP) was calculated for each prediction scenario as follows:

$$MSEP = \frac{1}{n} \sum_{i=1}^{n} (Y_i - \widehat{Y}_i)^2$$

where "*n*" is the number of individuals, "*Y*_i" is the observed phenotype for the ith individual, and " \hat{Y}_i " is the predicted value for the ith individual. Statistical analysis for each prediction was carried out using oneway ANOVA followed by multiple comparisons using Tukey's post hoc method. GraphPad Prism (version 8.0.2) software (GraphPad Software, Inc., San Diego, CA, USA) was used to plot graphs and statistical analysis. A *p*-value of < 0.05 was considered significant.

3. Results

3.1. Parentage assignment and genomic relatedness matrix

The parentage assignment of the 1009 offspring revealed that they belonged to 40 families. (Fig. 1 A). Of the total fish, both parents were assigned to 786 offspring, and maternal candidates were assigned to 861 fish. However, we could not assign any parental candidates to 148 of the 1009 fish. Approximately 37.76% of the population was established from two families that were maternal half-sibs (oP_1142). Based on the relatedness matrix, the total population was clustered into three sub-populations (Fig. 1B).

3.2. Predictive ability

After final QC filtering, 54,730 SNPs were retained from the 1009 successfully genotyped fish. These were used to estimate predictive ability by changing different factors.

3.2.1. Effect of prediction method and model

Eleven different prediction methods were used to estimate predictive ability. The two BLUP methods used in this study included GBLUP and EGBLUP. Five Bayesian methods (BA, BB, BC, BL, and BRR) and four ML methods (EN, RR, RKHS, and RF) were also used. According to Fig. 2 A and Table S1, all the methods showed predictive abilities ranging from 63.4% to 69.8%. Most methods used in this study did not significantly differ (p < 0.05) in their predictive ability. However, BB (0.675), EN (0.679), and RF (0.698) demonstrated significantly higher predictive abilities (p < 0.05). In addition, BB (0.164) and RF (0.162) showed the lowest MSEP compared with the other prediction methods. Therefore, GBLUP, BB, and RF from each category were used for further analyses.

Including different fixed effects in the model increased predictive

ability (Fig. 2B and Table S2). However, since the RF model cannot include fixed effects manually, we only used the BB and GBLUP models in this context. Including animal sex as a fixed effect significantly increased the model's predictive ability by 20%, and MSEP reduced for both GBLUP and BB. However, predictive ability did not change significantly with familial information.

3.2.2. Predictive ability by changing SNP marker number

As shown in Fig. 3 and Table S3, predictive abilities were affected by the number of SNPs around the QTL. When the SNP number increased from 100 to 1000, predictive ability significantly increased, whereas, at a lower marker number, we observed a significant deviation in predictive ability (Fig. 3A). GBLUP had the lowest predictive ability compared with the other prediction methods but gave stable predictions starting from 600 to 50,000 SNPs. BB and RF showed a plateau in predictive ability at larger SNP numbers (1000–50,000) and reduced standard deviation (SD; Fig. 3B and C). Notably, including sex as a fixed effect in GBLUP and BB showed that predictive ability did not considerably change with marker number.

3.2.3. Predictive ability by changing population size and fold

Predictive ability can be affected by the training and test population sizes. The small population size significantly reduced the predictive ability of the model (Fig. 4A and Table S4). A maximum of 1009 fish was used for the analysis. According to the results, the predictive ability of the mean models increased with increasing population size. In contrast, predictive ability was stable when more than 300 animals were used as the population size in the fixed effects models.

We then performed k-fold validation of the predictive ability of the models. According to the results (Fig. 4B and Table S5), the predictive abilities following 3- to 10-fold cross-validation did not significantly differ (p < 0.05) for all models.

3.2.4. Predictive ability by age of fish

A total of 682 fish with phenotype records (Table 1) and 54,463 QCpassed SNPs were used for prediction (Fig. 5 A and Table S6) and heritability estimation (Table 2). At the age of 1.2 years, the predictive ability for weight was reduced compared with that at other ages. In addition, at 1.8 years, predictive ability was significantly higher (p <0.05) for all models and methods. Furthermore, heritability of the growth traits increased with age.

3.2.5. Predictive ability for length and body depth

The predictive ability for the total length and body depth of 1.8-yearold fish was estimated using five prediction models (Fig. 5B and C). GBLUP_mean, BB_mean, and RF_mean models had predictive abilities for length ranging from 65.5% to 70.1%, and GBLUP_fix and BB_fix had predictive abilities of 85.7 and 85.2%, respectively (Table S7). For body depth prediction, GBLUP, BB, and RF exhibited predictive abilities of 66.5, 70.5, and 71.4%, respectively. GBLUP_fix and BB_fix exhibited high predictive abilities at 86.8 and 86.1%, respectively. The mean- and fixed-effect models significantly differed in predictive ability, which was similar to that observed for animal weight. Additionally, in the mean models, BB and RF showed significantly higher predictive abilities than GBLUP for both length and body depth.

4. Discussion

Our study determined the predictive ability for weight using different models and methods by changing different parameters. Initially, we checked the predictive abilities of 11 prediction methods: GBLUP, Bayesian, and ML. We used a basic mean model to test the different models without including any fixed effects. In fish, parameters such as sex, living temperature, and disease outbreaks can affect growth (Honeycutt et al., 2019; Imsland et al., 2019). When we are unable to measure these parameters, the mean model is used to get an idea of the



Fig. 1. (A) Parentage analysis of oFO_KW fish obtained using 4000 SNPs. (B). Heatmap of the genomic relatedness of oFO_KW fish.



Fig. 2. (A) Predictive ability for weight by the 11 prediction methods used in this study. (B) Predictive ability of models, including different fixed effects. *Y* is a vector of phenotypes, μ is the overall mean, *g* is a vector of random effects calculated using SNPs, *f* is the fixed effect of the family identified by parentage analysis, and *S* is the fixed effect of the sex of each fish. Different letters indicate the significance (p < 0.05) of predictive ability compared with that of the mean model of GBLUP or BB.

prediction. However, the predictive ability was lower than that of the optimal model with correctly identified parameters.

GBLUP is renowned for its simplicity and robustness and has been used for many genomic selection studies in plants and animals for years (Clark and van der Werf, 2013; de los Campos et al., 2013; Mouresan et al., 2019). It does not require computational complexity as in other methods, making it accessible for large-scale applications. In addition, GBLUP is well-suited for polygenic traits such as growth, as it assumes that all markers contribute to the genetic variance of the trait (Clark and van der Werf, 2013; Tiezzi and Maltecca, 2015) Correspondingly, we observed a moderate predictive ability of GBLUP, similar to other fish growth-related studies (Hosoya et al., 2021; Sukhavachana et al., 2021). However, its predictive ability was lower than that of BB and RF. This could be due to the limitations of GBLUP, which may be unable to capture the effect of rare variants with a larger impact on the trait because it assumes that all SNP markers contribute equally to the trait. In addition, this assumption does not provide insights into the biological mechanisms underlying trait variation. Other methods, such as Bayesian and ML, can potentially offer greater interpretability in this regard.

In our study, the predictive abilities of all the Bayesian methods and some ML methods, such as RR and RKHS, were not significantly different. However, of the Bayesian methods, BB showed a higher predictive ability than that of the other methods. BB encourages sparsity in the model, which assumes that only a subset of markers (usually a small proportion) has non-zero effects on the trait of interest. This leads to improved prediction by reducing the overfitting of large-scale genomic



Fig. 3. Predictive ability for weight when using (A) 100–1000 SNPs, (B) 1000–10,000 SNPs, and (C) 10,000–50,000 SNPs. Mean models and fixed effect (sex information)-included models of GBLUP and BB were used.

data, which could be the reason for the lower deviation and MSEP and higher predictive ability.

Among the ML methods, RF showed lower variance and higher predictive ability. ML methods are high-dimensional data-handling tools that can be useful in the present and future. They are easy to handle because of the automation of model building, learning, and accurate predictions (Chen and Ishwaran, 2012; Montesinos López et al., 2022). RF can handle nonlinear relationships between genotypic and phenotypic data and different interactions with genetic markers compared with ML methods (Bureau et al., 2003; Stephan et al., 2015; Wang et al., 2022). In olive flounder, growth is controlled by hormonal and environmental parameters, as well as genetics (Honeycutt et al., 2019; Ryu et al., 2020). Therefore, RF is well-suited for capturing these nonlinearities because it can build multiple decisions and combine their



Fig. 4. Predictive ability for weight by changing (A) population size and (B) k-fold.

Table 1

Crowth porformones	traits of the	study population	at each ago point
Growin periormance	traits of the	study population	at each age point.

Year (Age)	Sex	Average weight (g)	Average length (cm)	Average width (cm)
2020 June (1.2 years)	Total	760.31 ± 208.44	37.65±3.71	15.18±1.65
	Male	544.82 ± 108.33	$33.62{\pm}2.00$	$13.51 {\pm} 1.11$
	Female	841.119 ±177.19	39.14±3.01	$15.82{\pm}1.35$
2020 September	Total	1243.63 + 398.19	45.80±5.65	$17.99{\pm}2.33$
(1.5 years)	Male	740.54 +124.54	38.27±2.5	14.95±1.08
	Female	1432.29 +285.61	48.62±3.52	$19.13{\pm}1.50$
2021 January (1.8 years)	Total	1859.80 + 680.09	50.97±5.87	$20.82{\pm}3.02$
(2.2.) ((10))	Male	943.69 +168.62	42.92±2.35	16.61 ± 1.13
	Female	2203.34 +438.28	53.98±3.44	22.40±1.70
2021 September	Total	2139.43 +687.76	55.86±5.95	22.23±2.86
(2.3 years)	Male	1283.16 +257.99	47.94±2.85	18.57±2.02
	Female	2458.8 ±498.84	58.83±3.64	23.59±1.67

predictions. RR and RKHS are useful in this situation but are less effective than RF. Although RF can identify the genetic markers contributing most to prediction, RR and RKHS provide coefficient values for the important markers; therefore, they may not directly translate to the actual contribution of those markers (González-Camacho et al., 2018). In addition, RF is less prone to overfitting when dealing with

high-dimensional data containing a large number of genetic markers (Goldstein et al., 2011). RF is inherently robust for detecting outliers and noisy data points, which can critically affect model performance (Breiman, 2001). Therefore, RF may show a higher predictive ability for olive flounder weight than other mean model-derived methods.

Most ML models use both additive genetic effects and non-additive genetic effects to predict the gEBV. Although non-additive genetic effects are not transferred to subsequent generations, this approach can still improve predictive accuracy in several ways. This is because nonadditive genetic effects can capture the phenotypic variation, thus improving prediction accuracy by accounting for a larger proportion of genetic architecture underlying the trait (de Oliveira et al., 2023). Additionally, non-additive genetic effects contribute to the unfolding of complex traits (e.g., growth), which are driven by an interplay of various extrinsic and intrinsic factors. Specifically, some of these factors (e.g., epistasis), interact between genetic loci and play a crucial role in shaping phenotypic traits; however, these traits can be missed by models that only consider the additive effects (Alves et al., 2020). Therefore, incorporating non-additive genetic effects to the model in this study increased the accuracy of predicting breeding value and the genetic response of individuals (Onogi et al., 2021).

To include fixed effects, we used GBLUP and BB. By contrast, RF automatically builds the model; thus, we could not include fixed effects as outside parameters. Including parental information and other identified impactful parameters can significantly improve predictive ability (Sarinelli et al., 2019). Compared with GBLUP, BB showed a higher predictive ability for both fixed-effects-included and not-included models. Although including familial information did not significantly affect predictive ability. As expected, including fish sex in the model significantly increased the predictive ability of both methods. It is widely known that the growth of several fish, including olive flounder, depends on their sex, with female fish growing twice as much as males (Omeka et al., 2022). In our study, we observed an approximately 20% increase in predictive ability compared with that of the mean model when sex was included as a fixed effect in the model. Additionally, the predictive ability models which incorporated sex as a factor, exceeded 0.8. In selective breeding populations based on multiple families, a predictive ability exceeding 0.8 can be difficult. Few studies have shown that growth-related traits in fish such as tiger pufferfish and rainbow trout have higher predictive abilities in some models (Hosoya et al., 2021; Song and Hu, 2022). Predictive ability, often measured by metrics such as heritability or correlation coefficients, depends on various factors, including the genetic diversity within the population, accuracy of phenotypic measurements, selection criteria, and genetic architecture of the trait. In many species, there are differences between males and females in terms of phenotype and genetic expression. If these sexually dimorphic traits are under selection and have a high heritability, they can contribute to increased predictive ability. By segregating the data on the basis of sex, the model can more effectively capture the distinct genetic drivers of growth in each sex. This reduction in variability allows for more precise genomic predictions. Therefore, sex identification may be effective in both genomic and selective breeding.

However, one drawback in identifying fish sex is the lack of distinguishable morphological changes in juvenile male and female fish. Therefore, there is a need to opt for noninvasive, effective methods for sex identification. At mature ages, fish stripping can be used to identify sex; however, it takes 1–2 years for the fish to mature, considerably delaying the selection plan. In this context, the mean model can be used with an average prediction accuracy.

Our study revealed that the predictive ability was lower with a lower number of markers than with all markers for GBLUP and BB; at higher marker densities, predictive ability significantly increased by up to 14.3% and 20.3%, respectively, relative to that obtained with 100 SNPs. For 900–50,000 SNPs, predictive ability was higher but increased by 1.4% and 2.3% in GBLUP and BB, respectively. For RF, even for SNPs as low as 100, predictive ability was high, and increasing their number up



Fig. 5. (A) Predictive ability for weight with age. Predictive ability for the (B) length and (C) width of 1.8-year-old oFO_KW fish using selected models. Different letters indicate significant differences at p < 0.05.

Table 2Heritability of fish growth traits at each age point.

Year	Heritability weight	Heritability length	Heritability width
2020 June (1.2 years)	0.303	0.237	0.275
2020 September (1.5 years)	0.338	0.363	0.290
2021 January (1.8 years)	0.357	0.356	0.338
2021 September (2.3 years)	0.436	0.422	0.309

to 900 increased predictive ability by 7.6%. For 900–50,000 SNPs, predictive ability increased by 1.7%. In addition, the highest predictive ability for GBLUP, BB, and RF was observed for 4000, 3000–4000, and 4000–5000 SNPs, respectively. Other studies have also confirmed that 3000–5000 SNPs may be ideal for GP in fish (Hosoya et al., 2021; Song et al., 2023). In fixed effect-included models, predictive ability was greater than 80%, even for 100 SNPs. Increasing SNPs up to 50,000 increased predictive ability by 2.2%. Using more SNPs often increases predictive ability, enhances trait coverage, especially in polygenic traits, and captures rare variants that are not effectively identified by lower-density arrays. Conversely, increasing the number of SNPs can significantly increase genotyping costs and cause overfitting and noise when the sample size is limited (Meuwissen et al., 2001). In addition,

increasing SNP density improves predictive ability, but at some point, additional SNPs (diminishing returns) may not significantly improve prediction (Yang et al., 2010). Therefore, for future genomic selection of flounder, we could use a lower number of SNPs based on a method that could be beneficial in several ways.

Population size directly affects predictive ability, as described in previously (Takeda et al., 2021). Our data showed that increasing population size from 100 to 600 increased predictive ability by 16.6% for GBLUP, 22.2% for BB, and 9.9% for RF. Additionally, for a population size of up to 1000, GBLUP, BB, and RF showed increasing patterns of predictive ability and reduced SD. In the fixed effect-included GBLUP and BB models, predictive abilities tended to be more stable beyond population sizes of 400. Increasing the population size can increase predictive ability by increasing genetic diversity, improving marker effect estimation, reducing the impact of LD, providing greater statistical power, and reducing the variance of predictive ability (Daetwyler et al., 2010; Goddard, 2009; Heslot et al., 2012; Merrick and Carter, 2021; VanRaden, 2008). The optimal population size is case-dependent; however, a 3000-5000 population size is recommended for prediction (Haves et al., 2009). Conversely, larger population sizes can also lead to challenges, such as increased computational power requirements, data management complexities, and cost.

Typically, 5-fold cross-validation is used in GP studies for several reasons. It can improve the balance between over- and under-fitting. With fewer folds, such as 2- or 3-fold, the risk of overfitting is higher, where the predictive ability is good with reference data; nonetheless, it does not generalize well with new data (Schrauf et al., 2021). Therefore, using more folds, such as 10-fold, substantially increases the gains in accuracy; however, the computational and time costs can also increase. In addition, 5-fold cross-validation provides a reasonable compromise between a lower number of individuals for model training and reliable estimates of model performance. Splitting data into multiple folds can increase model performance by alleviating heterogeneity and variation due to different factors, such as genetic diversity, environmental influences, and experimental noise (Schrauf et al., 2021). Furthermore, it provides statistical consistency for GP, making it easier to compare results across other studies. Our data also revealed that 5-6-fold cross-validation reduces the MSEP.

The age of animals can also be an influential factor in the predictive ability of a model. In this study, the prediction accuracy increased from 1.2 to 1.8 years; however, at 2.8 years, a decrease in prediction accuracy was observed even at high heritability. Several factors can be affected by the changes in prediction accuracy over the age as flounder growth is controlled by hormones and genes activated during sexual maturation (Ryu et al., 2020). According to our data, weight gain in male and female flounders differ over time (Fig. S2). To be specific, female fish growth exponentially increased up to 1.8 years and then decreased at 2.3 years. In contrast, male fish, with a lesser growth rate, showed linear weight gain throughout the period. Therefore, changes in prediction accuracy might be caused by the growth curve differences. Few studies have been performed in cattle and boar to identify the effect of growth curve on the prediction accuracy of the model (Haraldsen et al., 2009; Yin and König, 2020). However, extensive studies need to determine which growth stage gives the highest prediction accuracy for flounder, and further studies on complex modeling for growth curve of olive flounder should be performed. In addition, age can be affected predictive ability owing to developmental changes characterized by physiological behavior, which change over age (Polverino et al., 2016).

Although fish length and body depth are not commercial parameters, they are highly phenotypically and genomically moderately correlated with fish weight (Omeka et al., 2022). Therefore, we checked the model performance for these traits using the same models as those used for weight prediction. We observed that length and body depth traits were also well predicted by the selected models. Therefore, when determining length and body depth parameters to collect more information, such as the condition factor, the same models can be used. The scope for future

studies might be using multi-trait prediction models weighted with important traits such as disease resistance and thermal tolerance. In addition, using a larger population with vast genetic diversity would be helpful in increasing the accuracy of the prediction. Moreover, the usage of the low SNP array should be assessed in depth in future studies as it significantly reduces the cost of the experiment.

In our study, the identified model parameters significantly affected model performance, and fine-tuning them can increase the predictive ability of the models. The target progeny can be used as brood stock in other genomic selection projects. Therefore, the results of this study will be valuable for the genomic selection of flounders in Korea.

5. Conclusion

This study estimated the GP accuracy of several models and the impact of different model components on olive flounder growth. RF showed the highest predictive ability in the mean model. Including sex as a fixed effect significantly increased the predictive ability of these methods. At lower marker numbers, predictive ability was highly variable; therefore, using more than 3000 markers could be effective. Increasing the population size and k-fold reduces the SD of the predictive ability. Overall, our results may be useful for implementing a gEBV-based selection strategy in olive flounder aquaculture.

CRediT authorship contribution statement

H.M.V. Udayantha: Investigation. Gaeun Kim: Investigation. D.S. Liyanage: Investigation, Methodology, Software, Visualization, Writing – review & editing. Sukkyoung Lee: Data curation, Investigation, Validation, Visualization. Jehee lee: Conceptualization, Funding acquisition, Investigation, Project administration, Resources, Supervision, Writing – review & editing. W.K.M Omeka: Conceptualization, Formal analysis, Investigation, Software, Visualization, Writing – original draft. Cecile Massault: Software, Supervision, Validation. Dean R. Jerry: Software, Supervision, Validation, Writing – review & editing. Taehyug Jeong: Formal analysis, Investigation. David B. Jones: Software, Validation, Writing – review & editing. Subothini Ganeshalingam: Investigation.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgments

This study was supported by the Basic Science Research Program of the National Research Foundation of Korea (NRF), funded by the Ministry of Education (2019R1A6A1A03033553) and the Korea Institute of Marine Science and Technology Promotion (KIMST), funded by the Ministry of Oceans and Fisheries (RS-2022-KS221670).

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.aqrep.2024.102132.

References

Alves, K., Luiz|, Brito, F., Baes, C.F., Sargolzaei, |Mehdi, Andrew, J., Robinson, B., Flavio, Schenkel, S, 2020. Estimation of additive and non-additive genetic effects for fertility

and reproduction traits in North American Holstein cattle using genomic information Funding information Agriculture and Agri-Food Canada; Dairy Farmers of Canada; Canadian Dairy Network; Canadian Dairy Commission. J. Anim. Breed. Genet 137, 316–330. https://doi.org/10.1111/jbg.12466.

Breiman, L., 2001. Random forests. Mach. Learn. 45, 5–32. https://doi.org/10.1023/A: 1010933404324.

- Bureau, A., Dupuis, J., Hayward, B., Falls, K., Van Eerdewegh, P., 2003. Mapping complex traits using Random Forests. BMC Genet 4, S64. https://doi.org/10.1186/ 1471-2156-4-S1-S64.
- Chaivichoo, P., Sukhavachana, S., Khumthong, R., Srisapoome, P., Chatchaiphan, S., Na-Nakorn, U., 2023. Genome–wide association study and genomic prediction of growth traits in bighead catfish (*Clarias macrocephalus* Günther, 1864). Aquaculture 562, 738748. https://doi.org/10.1016/J.AQUACULTURE.2022.738748.
- Chang, C.C., Chow, C.C., Tellier, L.C., Vattikuti, S., Purcell, S.M., Lee, J.J., 2015. Secondgeneration PLINK: rising to the challenge of larger and richer datasets. Gigascience 4, 7. https://doi.org/10.1186/S13742-015-0047-8.
- Chen, X., Ishwaran, H., 2012. Random forests for genomic data analysis. Genomics 99, 323–329. https://doi.org/10.1016/J.YGENO.2012.04.003.
- Clark, S.A., van der Werf, J., 2013. Genomic Best Linear Unbiased Prediction (gBLUP) for the Estimation of Genomic Breeding Values. in: Methods in Molecular Biology. Humana Press Inc, pp. 321–330. https://doi.org/10.1007/978-1-62703-447-0 13.
- D'Agaro, E., Favaro, A., Matiussi, S., Gibertoni, P.P., Esposito, S., 2021. Genomic selection in salmonids: new discoveries and future perspectives. Aquac. Int. 2021 295 29, 2259–2289. https://doi.org/10.1007/S10499-021-00747-W.
- Daetwyler, H.D., Pong-Wong, R., Villanueva, B., Woolliams, J.A., 2010. The impact of genetic architecture on genome-wide evaluation methods. Genetics 185, 1021–1031. https://doi.org/10.1534/genetics.110.116855.
- De Oliveira, L.F., Brito, L.F., Marques, D.B.D., da Silva, D.A., Lopes, P.S., Dos Santos, C. G., Johnson, J.S., Veroneze, R., 2023. Investigating the impact of non-additive genetic effects in the estimation of variance components and genomic predictions for heat tolerance and performance traits in crossbred and purebred pig populations. BMC Genom. Data 24, 1–18. https://doi.org/10.1186/S12863-023-01174-X/TABLES/8.
- Gianola, D., Van Kaam, J.B.C.H.M., 2008. Reproducing Kernel Hilbert spaces regression methods for genomic assisted prediction of quantitative traits. Genetics 178, 2289. https://doi.org/10.1534/GENETICS.107.084285.
- Goddard, M., 2009. Genomic selection: Prediction of accuracy and maximisation of long term response. Genetica 136, 245–257. https://doi.org/10.1007/S10709-008-9308-0/METRICS.
- Goldstein, B.A., Polley, E.C., Briggs, F.B.S., 2011. Random Forests for Genetic Association Studies. Stat. Appl. Genet. Mol. Biol. 10, 32. https://doi.org/10.2202/1544-6115.1691.
- González-Camacho, J.M., Ornella, L., Pérez-Rodríguez, P., Gianola, D., Dreisigacker, S., Crossa, J., 2018. Applications of Machine Learning Methods to Genomic Selection in Breeding Wheat for Rust Resistance. Plant Genome 11, 170104. https://doi.org/ 10.3835/plantgenome2017.11.0104.
- Haraldsen, M., Ødegård, J., Olsen, D., Vangen, O., Ranberg, I.M.A., Meuwissen, T.H.E., 2009. Prediction of genetic growth curves in pigs. animal 3, 475–481. https://doi. org/10.1017/S1751731108003807.
- Hayes, B.J., Bowman, P.J., Chamberlain, A.J., Goddard, M.E., 2009. Invited review: Genomic selection in dairy cattle: Progress and challenges. J. Dairy Sci. 92, 433–443. https://doi.org/10.3168/jds.2008-1646.
- Heslot, N., Yang, H.P., Sorrells, M.E., Jannink, J.L., 2012. Genomic Selection in Plant Breeding: A Comparison of Models. Crop Sci. 52, 146–160. https://doi.org/10.2135/ CROPSCI2011.06.0297.
- Honeycutt, J.L., Deck, C.A., Miller, S.C., Severance, M.E., Atkins, E.B., Luckenbach, J.A., Buckel, J.A., Daniels, H.V., Rice, J.A., Borski, R.J., Godwin, J., 2019. Warmer waters masculinize wild populations of a fish with temperature-dependent sex determination. Sci. Rep. 9, 6527. https://doi.org/10.1038/s41598-019-42944-x.
- Hosoya, S., Yoshikawa, S., Sato, M., Kikuchi, K., 2021. Genomic prediction for testes weight of the tiger pufferfish, *Takifugu rubripes*, using medium to low density SNPs. Sci. Rep. 11, 20372 https://doi.org/10.1038/s41598-021-99829-1.
- Imsland, A.K.D., Gunnarsson, S., Thorarensen, H., 2019. Impact of environmental factors on the growth and maturation of farmed Arctic charr. Rev. Aquac. 12, raq.12404 https://doi.org/10.1111/raq.12404.
- Jiang, Y., Reif, J.C., 2015. Modeling epistasis in genomic selection. Genetics 201, 759–768. https://doi.org/10.1534/genetics.115.177907.
- Jung, J.-Y., Kim, S., Kim, K., Lee, B.-J., Kim, K.-W., Han, H.-S., 2020. Feed and Disease at Olive Flounder (*Paralichthys olivaceus*) Farms in Korea. Fishes 5, 21. https://doi.org/ 10.3390/fishes5030021.
- Kang, J.-H., Noh, J.-K., Kim, J.-H., Lee, J.-H., Kim, H.-C., Kim, K.-K., Kim, B.-S., Lee, W.-J., 2006. Genetic relationship between broodstocks of olive flounder, *Paralichthys* olivaceus (Temminck and Schlegel) using microsatellite markers. Aquac. Res. 37, 701–707. https://doi.org/10.1111/j.1365-2109.2006.01483.x.
- Kim, J.-H., Lee, J.-H., Kim, H.C., Noh, J.K., Kang, J.-H., Kim, K.-K., 2011. Body Shape and Growth in Reciprocal Crosses of Wild and Farmed Olive Flounder, *Paralichthys olivaceus*. J. WORLD Aquac. Soc. 42 https://doi.org/10.1111/j.1749-7345.2011.00463.x.
- Li, Y., Zhang, B., Lu, S., Tian, Y., Yang, Y., Chen, S., 2018. Genetic parameters estimates for growth performance traits at harvest in Japanese flounder (*Paralichthys olivaceus*. Aquaculture 489, 56–61. https://doi.org/10.1016/j.aquaculture.2018.01.010.
- Li, Y., Zhang, B., Yang, Y., Chen, S., 2019. Estimation genetic parameters for juvenile growth performance traits in oliveflounder (*Paralichthys olivaceus*). Aquac. Fish. 4, 48–52. https://doi.org/10.1016/j.aaf.2018.12.001.
- Lin, Z., Hosoya, S., Sato, M., Mizuno, N., Kobayashi, Y., Itou, T., Kikuchi, K., 2020. Genomic selection for heterobothriosis resistance concurrent with body size in the

tiger pufferfish, Takifugu rubripes. Sci. Rep. 2020 101 10 (1), 13. https://doi.org/ 10.1038/s41598-020-77069-z.

- Liyanage, D.S., Lee, S., Yang, H., Lim, C., Omeka, W.K.M., Sandamalika, W.M.G., Udayantha, H.M.V., Kim, G., Ganeshalingam, S., Jeong, T., Oh, S.R., Won, S.H., Koh, H.B., Kim, M.K., Jones, D.B., Massault, C., Jerry, D.R., Lee, J., 2022. Genomewide association study of VHSV-resistance trait in *Paralichthys olivaceus*. Fish. Shellfish Immunol. 124, 391–400. https://doi.org/10.1016/J.FSL2022.04.021.
- de los Campos, G., Hickey, J.M., Pong-Wong, R., Daetwyler, H.D., Calus, M.P.L., 2013. Whole-Genome Regression and Prediction Methods Applied to Plant and Animal Breeding. Genetics 193, 327. https://doi.org/10.1534/GENETICS.112.143313.
- Lu, S., Liu, Y., Qu, S., Zhou, Q., Wang, L., Zhang, T., Xu, W., Zhang, M., Song, Y., Wang, J., Zhu, C., Chen, S., 2023. Genomic prediction of survival against Vibrio harveyi in leopard coral grouper (*Plectropomus leopardus*) using GBLUP, weighted GBLUP, and BayesCπ. Aquaculture 572, 739536. https://doi.org/10.1016/J. AQUACULTURE.2023.739536.
- Merrick, L.F., Carter, A.H., 2021. Comparison of genomic selection models for exploring predictive ability of complex traits in breeding programs. Plant Genome 14, e20158. https://doi.org/10.1002/tpg2.20158.
- Meuwissen, T.H.E., Hayes, B.J., Goddard, M.E., 2001. Prediction of total genetic value using genome-wide dense marker maps. Genetics 157, 1819. https://doi.org/ 10.1093/GENETICS/157.4.1819.
- Montesinos López, O.A., Montesinos López, A., Crossa, J., 2022. Random Forest for Genomic Prediction, in: Multivariate Statistical Machine Learning Methods for Genomic Prediction. Springer International Publishing, Cham, pp. 633–681. https:// doi.org/10.1007/978-3-030-89010-0_15.
- Mouresan, E.F., Selle, M., Rönnegård, L., 2019. Genomic Prediction Including SNP-Specific Variance Predictors. G3 Genes|Genomes|Genet. 9, 3333–3343. https://doi. org/10.1534/g3.119.400381.
- Ogutu, J.O., Schulz-Streeck, T., Piepho, H.P., 2012. Genomic selection using regularized linear regression models: Ridge regression, lasso, elastic net and their extensions. BMC Proc. 6, 1–6. https://doi.org/10.1186/1753-6561-6-S2-S10/TABLES/1.
- Omeka, W.K.M., Liyanage, D.S., Lee, S., Lim, C., Yang, H., Sandamalika, W.M.G., Udayantha, H.M.V., Kim, G., Ganeshalingam, S., Jeong, T., Oh, S.R., Won, S.H., Koh, H.B., Kim, M.K., Jones, D.B., Massault, C., Jerry, D.R., Lee, J., 2022. Genomewide association study (GWAS) of growth traits in olive flounder (*Paralichthys olivaceus*). Aquaculture 555, 738257. https://doi.org/10.1016/J. AQUACULTURE.2022.738257.
- Onogi, A., Watanabe, T., Ogino, A., Kurogi, K., Togashi, K., 2021. Genomic prediction with non-additive effects in beef cattle: stability of variance component and genetic effect estimates against population size. BMC Genom. 22, 1–9. https://doi.org/ 10.1186/S12864-021-07792-Y/FIGURES/4.
- Polverino, G., Cigliano, C., Nakayama, S., Mehner, T., 2016. Emergence and development of personality over the ontogeny of fish in absence of environmental stress factors. Behav. Ecol. Sociobiol. 70, 2027–2037. https://doi.org/10.1007/ s00265-016-2206-z.
- Ryu, J.-W., Jung, J., Park, K., Lee, S., Park, I., Kim, W.-J., Kim, D.-S., 2020. Characterization of sexual size dimorphism and sex-biased genes expression profile in the olive flounder. Mol. Biol. Rep. 47, 8317–8324. https://doi.org/10.1007/ s11033-020-05843-3.
- Sarinelli, J.M., Murphy, J.P., Tyagi, P., Holland, J.B., Johnson, J.W., Mergoum, M., Mason, R.E., Babar, A., Harrison, S., Sutton, R., Griffey, C.A., Brown-Guedira, G., 2019. Training population selection and use of fixed effects to optimize genomic predictions in a historical USA winter wheat panel. Theor. Appl. Genet. 132, 1247–1261. https://doi.org/10.1007/s00122-019-03276-6.
- Schrauf, M.F., de los Campos, G., Munilla, S., 2021. Comparing Genomic Prediction Models by Means of Cross Validation. Front. Plant Sci. 12, 734512 https://doi.org/ 10.3389/FPLS.2021.734512.
- Sohn, H., Kim, J., Jin, C., Lee, J., 2019. Identification of Vibrio species isolated from cultured olive flounder (*Paralichthys olivaceus*) in Jeju Island, South Korea. Fish. Aquat. Sci. 22, 14. https://doi.org/10.1186/s41240-019-0129-0.
- Song, H., Hu, H., 2022. Strategies to improve the accuracy and reduce costs of genomic prediction in aquaculture species. Evol. Appl. 15, 578–590. https://doi.org/ 10.1111/eva.13262.
- Song, H., Dong, T., Hu, M., Yan, X., Xu, S., Hu, H., 2022. First single-step genomic prediction and genome-wide association for body weight in Russian sturgeon (*Acipenser gueldenstaedtii*). Aquaculture 561, 738713. https://doi.org/10.1016/j. aquaculture.2022.738713.
- Song, H., Dong, T., Yan, X., Wang, W., Tian, Z., Sun, A., Dong, Y., Zhu, H., Hu, H., 2023. Genomic selection and its research progress in aquaculture breeding. Rev. Aquac. 15, 274–291. https://doi.org/10.1111/raq.12716.
- Stephan, J., Stegle, O., Beyer, A., 2015. A random forest approach to capture genetic effects in the presence of population structure. Nat. Commun. 6, 7432. https://doi. org/10.1038/ncomms8432.
- Sukhavachana, S., Senanan, W., Pattarapanyawong, N., Srithong, C., Joerakate, W., Tunkijjanukij, S., Poompuang, S., 2021. Multiple-trait genomic prediction of harvest and fillet traits in Asian seabass (*Lates calcarifer*, Bloch 1790). Aquaculture 544, 737069. https://doi.org/10.1016/j.aquaculture.2021.737069.
- Takeda, M., Inoue, Keiichi, Oyama, H., Uchiyama, K., Yoshinari, K., Sasago, N., Kojima, T., Kashima, M., Suzuki, H., Kamata, T., Kumagai, M., Takasugi, W., Aonuma, T., Soma, Y., Konno, S., Saito, T., Ishida, M., Muraki, E., Inoue, Y., Takayama, M., Nariai, S., Hideshima, R., Nakamura, R., Nishikawa, S., Kobayashi, H., Shibata, E., Yamamoto, K., Yoshimura, K., Matsuda, H., Inoue, T., Fujita, A., Terayama, S., Inoue, Kazuya, Morita, S., Nakashima, R., Suezawa, R., Hanamure, T., Zoda, A., Uemoto, Y., 2021. Exploring the size of reference population for expected accuracy of genomic prediction using simulated and real data in

W.K.M. Omeka et al.

Japanese Black cattle. BMC Genom. 22, 799. https://doi.org/10.1186/s12864-021-08121-z.

- Tiezzi, F., Maltecca, C., 2015. Accounting for trait architecture in genomic predictions of US Holstein cattle using a weighted realized relationship matrix. Genet. Sel. Evol. 47, 24. https://doi.org/10.1186/s12711-015-0100-1.
- Udayantha, H.M.V., Lee, S., Liyanage, D.S., Lim, C., Jeong, T., Omeka, W.K.M., Yang, H., Kim, G., Kim, J., Lee, Jihun, Nadarajapillai, K., Ganeshalingam, S., Park, C.U., Lee, Jiwon, Oh, S.R., Gong, P., Jang, Y., Hyun, J., Park, A., Koh, H.B., Kim, M.K., Jones, D.B., Massault, C., Jerry, D.R., Lee, Jehee, 2023. Identification of candidate variants and genes associated with temperature tolerance in olive flounders by Genome-Wide Association Study (GWAS). Aquaculture 576. https://doi.org/ 10.1016/J.AQUACULTURE.2023.739858.
- VanRaden, P.M., 2008. Efficient Methods to Compute Genomic Predictions. J. Dairy Sci. 91, 4414–4423. https://doi.org/10.3168/jds.2007-0980.
- Verbyla, K.L., Kube, P.D., Evans, B.S., 2022. Commercial implementation of genomic selection in Tasmanian Atlantic salmon: Scheme evolution and validation. Evol. Appl. 15, 631–644. https://doi.org/10.1111/eva.13304.

- Wang, K., Yang, B., Li, Q., Liu, S., 2022. Systematic Evaluation of Genomic Prediction Algorithms for Genomic Prediction and Breeding of Aquatic Animals. Genes (Basel) 13, 2247. https://doi.org/10.3390/genes13122247.
- Wang, T., Wu, X., Song, L., Yang, Y., Gong, S., Zeng, L., Tao, Y., Zhong, C., Meng, Z., Liu, X., 2023. Identification of candidate growth-related SNPs and genes using GWAS and transcriptome analyses in leopard coral grouper (Plectropomus leopardus). https://doi.org/10.1016/j.aquaculture.2023.739677.
- Yang, J., Benyamin, B., McEvoy, B.P., Gordon, S., Henders, A.K., Nyholt, D.R., Madden, P.A., Heath, A.C., Martin, N.G., Montgomery, G.W., Goddard, M.E., Visscher, P.M., 2010. Common SNPs explain a large proportion of the heritability for human height. Nat. Genet. 42, 565–569. https://doi.org/10.1038/ng.608.
- Yin, T., König, S., 2020. Genomic predictions of growth curves in Holstein dairy cattle based on parameter estimates from nonlinear models combined with different kernel functions. J. Dairy Sci. 103, 7222–7237. https://doi.org/10.3168/jds.2019-18010.
- Zou, H., Hastie, T., 2005. Regularization and variable selection via the elastic net. J. R. Stat. Soc. B 67, 301–320.