

582. Prediction of production traits by using body features of gilthead seabream (*Sparus aurata*) obtained from digital images

B. Gulzari^{1*}, A. Mencarelli², C. Roozeboom¹, H. Komen¹ and J.W.M. Bastiaansen¹

¹Animal Breeding and Genomics Group, Wageningen University & Research, the Netherlands;

²Wageningen Plant Research, Wageningen University & Research, the Netherlands; benan.gulzari@wur.nl

Abstract

Gilthead seabream (*Sparus aurata*) is a key aquaculture species in the Mediterranean and surrounding regions. While many traits are of interest in seabream breeding programs, phenotypes of them cannot always be easily or accurately measured. Therefore, the objectives are to predict phenotypes of production traits by using automated measurements of body features from digital images and to obtain genetic correlations between measured and predicted phenotypes. The production traits analysed were harvest weight (HW), fillet weight (FW), and fillet percentage (F%). Each image feature was tested for prediction of phenotypes by using 10-fold cross-validation. The phenotypic and genetic correlations of measured and predicted phenotypes were 0.98 and 0.996 for HW, 0.93 and 0.99 for FW, 0.27 and 0.70 for F%, respectively. The relative efficiency of mass selection was higher for predicted phenotypes, except for HW.

Introduction

Gilthead seabream (*Sparus aurata*) is a key aquaculture species in the Mediterranean and surrounding regions. The aquacultural production of gilthead seabream increased by 90% between 2009 and 2019 to 260,000 tons across 20 countries (FAO, 2021). However, only 60% of farmed seabream originated from commercial breeding programs in 2015 (Janssen *et al.*, 2015). In fish breeding programs, various traits are selected for to increase the production level, efficiency, and welfare of the fish. These traits are important for farmers as well as for other stakeholders. However, phenotypes of them cannot always be easily or accurately measured. Some traits, such as product quality traits, cannot be measured on live fish. Sib information can be used to select for such traits but using phenotypes of selection candidates is preferable and has three main advantages. First, both within family and between family variation can be utilized. Second, the accuracy of the estimated breeding values (EBV) will be higher. Third, using phenotypes of the selection candidates simplifies the breeding program and reduces the costs because fewer animals are needed to reach a certain accuracy of the EBVs.

Instead of actual phenotypes of difficult to measure traits, predictions of those phenotypes can be used for selection candidates. Predictions of the phenotypes from body measurements has been studied in various fish species. Manual measurements work well to predict the phenotypes of difficult to measure traits; however, manually measuring body features is labour intensive. Computer vision allows digital image acquisition and image processing for measurements of the fish by using dedicated software (Dowlati *et al.*, 2012). The measurements obtained from the images of fish can be used to establish predictions of the phenotypes. Furthermore, obtaining measurements from digital images can be automated (Navarro *et al.*, 2016).

Therefore, the objectives are to predict phenotypes of production traits by using automated measurements of body features from digital images and to obtain genetic correlations between measured and predicted phenotypes.

Materials & methods

Phenotypic data collection. For this experiment, a population of juveniles was produced by mass spawning of 33 males and 20 females on a single day. When fish reached an average weight of 3 grams, a batch of juveniles was stocked in a sea cage in southeast of Spain and grown under commercial conditions. Data on production traits were collected from the commercially produced fish that were harvested after a grow-out period of 500 days. After measuring harvest weight (HW), the fish were filleted, and fillet weight was recorded. Fillet percentage (F%) was calculated as $(FW / HW) \times 100$.

Each fish was photographed after being taken out of the water. Measurements of body features from digital images were automated by using the Halcon 18.11 framework (MV Tec Software GmbH, Munich, Germany). The body features measured from images included total area, fillet area, trunk area, head area, tail area, total length, standard length, height, head height, head length, total volume, and head excluded volume. Head proportion was calculated as head area divided by total area.

Prediction of traits. All variables measured on images were considered individually and combined as candidate predictors for each phenotype. Prediction accuracy was defined as the correlation between measured and predicted phenotypes and tested by using 10-fold cross-validation. Cross-validation was repeated 10 times to treat all the parts of the data as validation and test sets. Predictors were added to linear regression models with forward selection. For genetic analyses, each phenotype was predicted by using a linear model with all data and the combination of predictors that yielded the highest accuracy in cross-validation.

Estimation of heritabilities and correlations. Genetic correlations (r_g) between measured and predicted phenotypes were estimated by using a bivariate animal model in ASReml version 4.1 (Gilmour *et al.*, 2015). The bivariate animal model is:

$$y = X\beta + Zu + e \quad (1)$$

where y is a vector of measured and predicted phenotypes, β is the vector of fixed effect 'sampling day', u is the vector of random animal additive genetic effects:

$$\sim \left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, G \begin{bmatrix} \sigma_{a,T_M}^2 & r_{a,T_M,P} \sigma_{a,T_M} \sigma_{a,T_P} \\ r_{a,T_M,P} \sigma_{a,T_M} \sigma_{a,T_P} & \sigma_{a,T_P}^2 \end{bmatrix} \right),$$

where G is the genomic relationship matrix and σ_{a,T_M}^2 is the additive genetic variance of measured trait, σ_{a,T_P}^2 is the additive genetic variance of predicted trait, $r_{a,T_M,P}$ is the additive genetic correlation between the measured and predicted traits, and e is the vector of random residual effects

In case of mass selection, the relative efficiency (RE) was calculated as:

$$RE = \left(\sqrt{h_p^2 / h_M^2} \right) \times r_A \quad (2)$$

where h_p^2 is the heritability of predicted phenotypes, h_M^2 is the heritability of measured phenotypes, and r_A is the genetic correlation between measured and predicted phenotypes.

Results

Prediction of harvest weight. The accuracy of predicting phenotypes was 0.98 for HW, and the genetic correlation between measured and predicted phenotypes was 0.996 ± 0.001 . The model that yielded the highest accuracy of prediction contained head excluded volume and total area of the fish. The heritability

of the predicted HW (0.52 ± 0.05) was slightly lower than measured HW (0.54 ± 0.05). The RE of indirect selection for HW was 98%.

Prediction of fillet weight and fillet percentage. The accuracy of predicting phenotypes was 0.92 and 0.27 for FW and F%, and the genetic correlation between measured and predicted phenotypes was 0.99 ± 0.005 and 0.74 ± 0.13 . The model that yielded the highest correlation for FW contained head excluded volume and total area of the fish. The heritability of the predicted FW (0.52 ± 0.05) was slightly higher than measured FW (0.47 ± 0.05). The relative efficiency of indirect selection for FW was 104%. The model that yielded the highest correlation for F% contained head proportion and head length of the fish. The heritability of the predicted F% (0.31 ± 0.05) was higher than measured F% (0.12 ± 0.04). The RE of indirect selection for F% was 119%.

Discussion

Our findings show that predicting phenotypes of gilthead seabream from digital images has strong potential to replace manual phenotyping in breeding programs. In case of mass selection, changing to predicted phenotypes would yield 4 and 19% more response for FW and F%, respectively. For HW, using predicted phenotypes would yield only 2% less response than using measured phenotypes.

Automation of phenotyping for fillet traits with digital images is advantageous due to two interrelated reasons. First reason is that the labour required for manual phenotyping increases if the breeding goal includes more than HW, especially when traits cannot be measured on live fish and siblings need to be phenotyped. In such cases, using image features to predict the phenotypes simplifies the breeding program because separate measurements for different traits will no longer be necessary and the fish does not have to be slaughtered. With a single image of the fish, it is possible to predict phenotypes for many traits. The second reason is that phenotypes can be obtained for selection candidates. Our findings imply that using digital images to phenotype for fillet traits is not only possible but also preferred in fish breeding programs. If predicted phenotypes are used, more response would be obtained in fillet traits and the response in HW would only slightly reduce in comparison to using measured phenotypes. Prediction of phenotypes provides benefits in genomic selection schemes as well. Predicted phenotypes become available on the selection candidates and the reference population contributes images and measured phenotypes to the training dataset for the genomic selection as well as the phenotype prediction models. Furthermore, prediction of phenotypes from images could enable application of genomic selection without genotyping the reference population. Genomic EBVs can be calculated for the selection candidates with the phenotypes predicted from image features. Not genotyping the reference population will affect the accuracy of genomic selection. This trade-off between the accuracy of genomic selection and accuracy of predicting phenotypes needs to be further investigated and optimized for each genomic selection scheme.

Certain strengths and limitations were recognized in this study. The main strengths of this study are that: (1) the measurements from digital images were automated, so the measurement error is minimal; (2) a genomic relationship matrix was used to estimate the genetic parameters. The advantage of using a genomic relationship matrix is that it improves the accuracy of the parameter estimates in comparison to using pedigree relationships (Veerkamp *et al.*, 2011). This study was performed with images from a single experimental seabream population that was produced as one batch with a full sib and half sib family structure. Accuracy of predicted phenotypes may reduce across different production batches or in unrelated populations of seabream. To avoid effects of overfitting, the predictions of phenotypes were evaluated with cross-validation, which is expected to result in a model with the strongest prediction ability to other populations. Before applying these models to predict the phenotypes of other seabream populations, it will be needed to test the model accuracies obtained in those populations. Extending the current dataset with

other unrelated seabream populations may change the accuracy of the predictions. Similar to reference datasets in genomic selection, the level of prediction accuracy will need to be assessed for situations where reference phenotype and image data are from more diverse background or when they are more distant from selection candidates.

Funding

This study is part of the AquaIMPACT project and was funded by the European Union's Horizon 2020 research and innovation programme under grant agreement No 818367.

References

- Dowlati, M., de la Guardia, M., Dowlati, M., Mohtasebi, S.S., 2012. Application of machine-vision techniques to fish-quality assessment. *TrAC Trends Anal. Chem.* 40, 168–179. <https://doi.org/10.1016/J.TRAC.2012.07.011>
- FAO, 2021. Fishery and Aquaculture Statistics. Global aquaculture production 1950-2019 (Fishstat).
- Gilmour, A.R., Gogel, B.J., Cullis, B.R., Welham, S.J., Thompson, R., 2015. ASReml user guide release 4.1 structural specification. Hemel Hempstead VSN Int. ltd.
- Janssen, K., Chavanne, H., Berentsen, P., Komen, H., 2015. Gilthead Seabream (*Sparus aurata*)—Current status of selective breeding in Europe, in: Abstract Book of the International Symposium on Genetics in Aquaculture XII, Santiago de Compostela, Spain.
- Navarro, A., Lee-Montero, I., Santana, D., Henriquez, P., Ferrer, M.A., Morales, A., Soula, M., Badilla, R., Negrín-Báez, D., Zamorano, M.J., Afonso, J.M., 2016. IMAFISH_ML: A fully-automated image analysis software for assessing fish morphometric traits on gilthead seabream (*Sparus aurata* L.), meagre (*Argyrosomus regius*) and red porgy (*Pagrus pagrus*). *Comput. Electron. Agric.* 121, 66–73. <https://doi.org/10.1016/j.compag.2015.11.015>
- Veerkamp, R.F., Mulder, H.A., Thompson, R., Calus, M.P.L., 2011. Genomic and pedigree-based genetic parameters for scarcely recorded traits when some animals are genotyped. *J. Dairy Sci.* 94, 4189–4197. <https://doi.org/10.3168/jds.2011-4223>