All fish deserve a breeding program

Designing affordable genomic selection programs with automated image analysis

Benan Gulzari

Benan Gulzari All fish deserve a breeding program

Propositions

1. Automated phenotyping is the master key to unlock the full potential of integrated breeding programs.
   (this thesis)
2. Image analysis removes the need to have a reference population in genomic selection designs.
   (this thesis)
3. For a successful PhD study, it is necessary to regularly meet with yourself.
4. Accessing scientific papers through online search engines lessens the importance of publishing in highly reputable scientific journals.
5. Imagination and innovation will be humans' most valuable assets in the era of artificial intelligence.
6. The high expectations from insect protein will only be realized if insects are used exclusively as fish feed.

Propositions belonging to the thesis, entitled

All fish deserve a breeding program - designing affordable genomic selection programs with automated image analysis

Benan Gulzari

Wageningen, 2 October 2023

All fish deserve a breeding program

designing affordable genomic selection programs with automated image analysis

Benan Gulzari

Thesis committee

Promotor

Prof. Dr. Hans Komen

Professor of Animal Breeding and Genomics

Wageningen University & Research

Co-promotor

Dr. John W.M. Bastiaansen

Researcher Animal Breeding and Genomics

Wageningen University & Research

Other members

Prof. Dr. Florian Muijres, Wageningen University & Research, the Netherlands

Prof. Dr. Hans Magnus Gjøen, Norwegian University of Life Sciences, Norway

Dr. Marco Bink, Hendrix Genetics, the Netherlands

Dr. Clemence Fraslin, The University of Edinburg, United Kingdom

This research was conducted under the auspices of the Graduate School of Wageningen Institute of Animal Sciences (WIAS).

All fish deserve a breeding program

designing affordable genomic selection programs with automated image analysis

Benan Gulzari

Thesis

Submitted in fulfilment of the requirements for the degree of doctor

at Wageningen University

by the authority of the Rector Magnificus

Prof. Dr. A.P.J. Mol,

in the presence of the

Thesis Committee appointed by the Academic Board

to be defended in public

on Monday 2 October 2023

at 16:00 in Omnia.

Benan Gulzari

All fish deserve a breeding program – Designing affordable genomic selection programs with automated image analysis

172 pages

PhD thesis, Wageningen University, Wageningen, the Netherlands (2023)

With references, with summary in English

ISBN: 978-94-6447-826-6

DOI: https://doi.org/10.18174/635785

Abstract

Gulzari, B. (2023). All fish deserve a breeding program – designing affordable genomic selection programs with automated image analysis. PhD thesis, Wageningen University & Research, the Netherlands.

Only a small fraction of the farmed fish species is genetically improved. Consequently, the majority of the farmed fish species are unable to provide the benefits of genetic improvement. Systematic genetic improvement is achieved within breeding programs, which can be distinguished as central-nucleus and integrated breeding programs. In central-nucleus breeding program designs, the broodstock and selection candidates are typically maintained in bio-secure indoor facilities, where environmental conditions can be controlled to be optimal and free of specific pathogens. Integrated breeding programs are approaches for genetic improvement of fish in which data collection and selection is performed in the commercial production environment. Due to the low cost of implementation, the integrated approach is well suited for fish species that are farmed at a relatively small scale. The main advantage of integrated breeding programs is that farmers have direct control over the breeding program and can select the animals that are adapted to their specific needs and conditions. In addition, the dissemination step is simply transferring the genetically improved material to on-growing facilities within the same farm. Breeding programs are complex operations and there are several challenges associated with their implementation. In this PhD thesis, I aim to contribute to the improvement of integrated breeding programs for fish species with innovative research. First, I addressed the challenge of obtaining phenotypes fast and accurately. This was achieved with a novel automated image analysis process for the extraction of morphometric measurements from 2D and 3D images of gilthead seabream. Second, I addressed the challenge of controlling the costs of breeding programs with a novel genomic selection design that eliminates the need to genotype a reference population. Third, I addressed the challenge of optimizing logistics and costs related to phenotyping by investigating the balance between the selection response in harvest weight and a difficult-to-measure trait in a two-stage integrated genomic selection design. Fourth, I estimated GxE for production traits between distinct production locations of gilthead seabream to decide if data from several environments should be incorporated into the breeding program. Unifying the knowledge produced in this thesis leads to improved implementation of integrated breeding programs to achieve tailored-to-the-location and affordable genetic improvement for fish species.

Contents

Chapter 1:

General Introduction

1.1 Genetic improvement of fish species

More than 500 fish species were farmed globally in 2016 (FAO, 2021). Commercial production of these fish is largely based on broodstock collected from the wild (Gjedrem, 2005). Wild stocks that are kept under commercial conditions are often stressed and unable to thrive (Gjedrem, 2000). Selective breeding is a well-known process to create animals that are better adapted to commercial production conditions (Gjedrem and Robinson, 2014). The benefits of genetic improvement are well-known and include increased production and yield (García-Ballesteros et al., 2022; Næve et al., 2022), improved utilization of resources (Kause et al., 2022), higher disease resistance and survival (Gulzari, 2017; Megahed, 2020) and enhanced animal welfare (Gharbi et al., 2015; Grimsrud et al., 2013). Systematic genetic improvement in fish started relatively recently for Atlantic salmon (*Salmo salar*), in 1975 (Gjedrem, 2010). Since then, breeding programs have been initiated for at least 30 fish species (Teletchea and Fontaine, 2014). Naturally, the degree of genetic improvement differs between those species and programs.

In general, breeding programs are considered complex and expensive, requiring resources and long-term commitment. This hampers the adoption of selective breeding for a high number of fish species that are currently cultured. In this thesis, I will investigate methods to make breeding programs simpler, more affordable, and easier to implement.

1.2 Breeding program designs

1.2.1 Central nucleus breeding programs

Central nucleus breeding program design is a common approach for systematic genetic improvement of fish species. In this approach, the broodstock and selection candidates are typically maintained in bio-secure indoor facilities, where environmental conditions can be controlled to be optimal and free of specific pathogens (Sang et al., 2020). The genetically improved material is then disseminated to fish farms from this “central” location (Gjedrem, 2010). Multipliers may be used if there is a large market for genetically improved material, the

species in question is not fecund enough to disseminate from a single location, or the capacity of the nucleus is insufficient to supply enough fry. Dissemination from central nucleus breeding programs may require comprehensive arrangements and management between several institutions (Tayamen, 2004), which can cause a time lag for genetically improved material to reach the fish farms. The flow of the animals and information in a central nucleus breeding program is depicted in Figure 1.1.

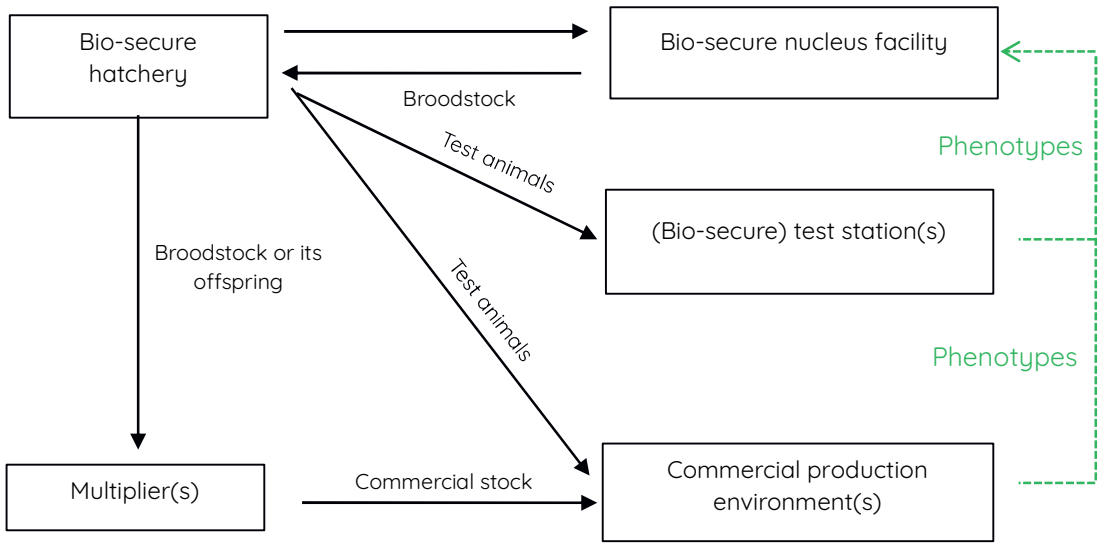

Figure 1.1. The flow of animals (black arrows) and information (green arrows) in central nucleus breeding programs.

In central nucleus breeding programs, the candidates' performances are measured in the central location with performance of relatives in test stations incorporated in a selection index. Test stations are used to phenotype for difficult-to-measure traits on siblings of the candidates, such as slaughterhouses for carcass traits (Prchal et al., 2020), containment facilities for disease challenge tests (Ødegård et al., 2006), or commercial production environments for production traits (Yoshida et al., 2022). Central-nucleus breeding programs are commonly used by breeding companies whose business is mainly aimed at selling genetically improved material (eggs or fingerlings) rather than profiting

directly from harvested fish or fish products. Therefore, central nucleus approach is suited more to fish species in high demand, such as Atlantic salmon (Næve et al., 2022), and Nile tilapia (Ponzoni et al., 2007), so that selling genetically improved material to fish farms would return enough profit for the breeding company.

1.2.2 Integrated breeding programs

Integrated breeding programs are approaches for genetic improvement of fish in which data collection and selection is performed in the commercial production environment. In this approach, there is a closed flow of genetic material in a fish farm (Figure 1.2) as the selected fish from the production environment are transferred to the hatchery as broodstock to produce offspring for production.

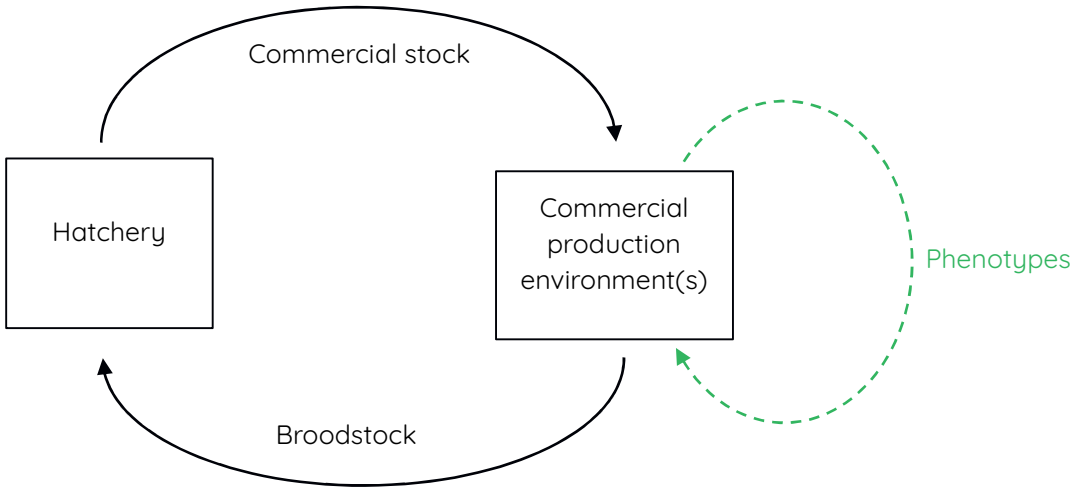

Figure 1.2. The flow of animals (black arrows) and information (green arrows) in integrated breeding programs.

Therefore, genetic improvement is integrated with commercial production within a fish farm. In this approach, harvested fish generates the main source of income rather than selling genetically improved material. In integrated breeding

programs, the cost of producing selection candidates is blended with the costs of producing the commercial stock. The extra hatchery and grow-out costs to produce the selection candidates are very small for fish farming companies that are using their own hatcheries to produce their own commercial stock.

Integrated breeding programs offer various advantages. Due to the low cost of implementation, the integrated approach is well suited for fish species that are farmed at a relatively small scale, for which a central-nucleus approach may not be profitable. The main advantage of integrated breeding programs is that farmers have direct control over the breeding program and can select the animals that are adapted to their specific needs and conditions. In addition, the dissemination step is simply transferring the genetically improved material to on-growing facilities within the same farm. A disadvantage of integrated approach is that the selection candidates and broodstock are exposed to natural threats such as diseases, environmental disruptions, and predators, which may interrupt or even destroy the breeding program.

1.3 Improvement of integrated fish breeding programs

Despite being simple in design, integrated fish breeding programs are complex operations and there are several challenges associated with their implementation. These challenges include obtaining phenotypes in a fast and accurate way, improving the accuracy of breeding value estimations, and neutralizing the undesired effects of GxE.

Phenotyping is an important challenge, which can be a financial and logistical bottleneck for integrated fish breeding programs when data must be collected on thousands of animals from the production environment. Ideally, phenotyping should be fast, easy, and accurate. In recent years, technological developments have improved phenotyping methods with sensor technology and computer vision (Saberioon et al., 2017). Computer vision simplifies and accelerates extracting many morphometric measurements simultaneously via image capture and subsequent extraction of information from digital images (Dowlati

et al., 2012). In recent years, automated image analysis methods have been developed, *e.g.,* (Navarro et al., 2016), that yield several morphometric measurements in seconds from a single image of the fish. However, challenges remain to use image analysis efficiently to obtain phenotypes of carcass traits.

Another important challenge in integrated breeding programs is to accurately rank the selection candidates. The simplest method to rank the candidates is using phenotypes, which can only be used to improve traits measured on live animals while accuracy is limited by heritability. Although this ranking method is primitive, some fish breeding programs still use it due to its simplicity (Chavanne et al., 2016; K Janssen et al., 2017). A more sophisticated way of ranking selection candidates is using the phenotypes of relatives in addition to own phenotypes of selection candidates, which together contribute to breeding value estimations with the use of a pedigree. This allows selection for difficult-to-measure traits measured on the relatives of selection candidates. In addition, the accuracies can be increased by manipulating the number of relatives with phenotypes. Depending on the heritability, high accuracies of breeding values can be obtained.

Estimation of breeding values using genetic markers has been receiving increased attention in recent years (Symonds et al., 2019). Genomic selection is an advanced implementation of breeding value estimation that accelerates genetic gain compared to pedigree-based selection methods (Meuwissen et al., 2001). This effect of genomic selection is mainly due to improved accuracy of estimated breeding values from utilizing all within-family genetic variation (Vallejo et al., 2017). In terrestrial species, genomic selection further accelerates genetic gain by reducing the generation interval (García-Ruiz et al., 2016). In fish breeding programs, there is limited opportunity to further reduce the generation interval because in most species, it is rather short and selection is performed before sexual maturity in most fish species (Boudry et al., 2021). Despite the benefits of genomic selection, high costs of genotyping impede the adoption of

this technology for fish species that are produced at a smaller scale than the species that are in high demand. Much research has focused on reducing genotyping costs including lowering the density of SNP panels (Kriaridou et al., 2020), imputation of SNPs (Tsai et al., 2017), and reusing a single reference population in several generations (Sonesson and Meuwissen, 2009). Selective genotyping strategies based on performance have also been studied to reduce the genotyping costs (Chu et al., 2020). In conclusion, genomic selection is beneficial in terms of improved selection response; however, widespread implementation of this technology in fish breeding programs remains a challenge.

The main aim of breeding programs is to create high performing populations under commercial conditions and GxE can diminish the effectiveness of breeding programs. In integrated breeding programs, GxE can be an issue if a farm produces fish in environmentally distinct locations. GxE occurs if the relative performance of genotypes changes in response to variations in environmental conditions (Falconer and Mackay, 1996). Reranking of genotypes is an undesired result of strong GxE, in which case the best genotypes in one condition may not be the best in other conditions. In the presence of GxE, only a part of the genetic gain can be commercially realized if the production environment is different than the selection environment (such as a bio-secure environment). This effect has been reported in aquaculture for growth of *Penaeus monodon* between a bio-secure recirculating system (nucleus environment) and outdoor grow-out ponds (Sang et al., 2020). In addition, GxE has been reported across varying oxygenation conditions (Mengistu et al., 2020), salinity levels (Domingos et al., 2021), and nutrition conditions (Romana-Eguia and Doyle, 1992). In relation to GxE, commercial fish production also faces challenges because of climate change. Water temperature is an important factor for fish growth (Mayer et al., 2012), which has been identified as one of the most important variables responsible for GxE in rainbow trout (Sae-Lim et al., 2014). In the Mediterranean,

the climate change unevenly affected the sea surface water temperature in different regions (Pastor et al., 2018), which can become a major source of uncertainty for breeding programs because the top performing genotypes in current temperature profile may not be the best in upcoming years.

1.4 Objectives and outline of this thesis

In this PhD thesis, I aim to contribute to the improvement of integrated breeding programs for fish species with innovative research. This thesis includes four research chapters, each of which addresses a different challenge faced by integrated breeding programs. In those chapters, I used gilthead seabream (*Sparus aurata*) as a case study to advance automated high-throughput phenotyping methods, optimize the allocation of phenotyping effort, improve the accuracy of estimated breeding values, and investigate GxE between production systems.

My objectives in **Chapter 2** are to develop a novel automated process for the extraction of morphometric measurements from 2D and 3D images of gilthead seabream, to estimate their genetic parameters, in particular their genetic correlations with the traits of interest for aquaculture production, and to validate their utility as indicator traits. An automated phenotyping process simplifies data collection on large numbers of fish and selecting on indicator traits may boost genetic progress because it allows exploiting Mendelian sampling variance.

My objectives in **Chapter 3** are to improve the prediction accuracy for difficult-to-measure traits using morphometric measurements in gilthead seabream that were obtained in Chapter 2, estimate the genetic correlations between observed and predicted phenotypes, and compare the accuracy of GEBVs between standard genomic selection designs and a novel design where genetic analyses are performed using predicted phenotypes of selection candidates. The novel design integrates automated high-throughput phenotyping with genomic selection, which may contribute to reducing the genotyping costs.

In **Chapter 4**, I investigated the optimal division of animals into selection candidates and reference population to obtain the desired selection responses for harvest weight or a difficult-to-measure trait under constraint genotyping efforts. For this, I simulate a two-stage selection scheme in which varying proportions are preselected on harvest weight and the candidates that are not preselected are slaughtered to obtain data on a difficult-to-measure trait. Depending on the priorities of fish farms, this scheme can balance the phenotyping effort for desired selection responses in harvest weight and a difficult-to-measure trait.

My objective in **Chapter 5** is to quantify GxE for production traits and organ weights of gilthead seabream in two distinct commercial production sites. For this, I used data collected on commercially produced seabream and investigated the degree of GxE between distinct production locations to determine if separate breeding programs are necessary for those environments. Understanding the degree of GxE also enables evaluation in terms of expected performance of genetically improved material in distinct locations.

In **Chapter 6 (General Discussion)**, I provide a synthesis of the knowledge produced in chapters 2-5 and how this knowledge can be used to improve integrated breeding programs for simple and affordable genetic improvement of all fish species.

Chapter 2:

High Quality Indicator Traits for Residual Fillet Weight and Fillet Fat% in Gilthead Seabream (*Sparus Aurata)* with 2D – 3D Image Analysis

Benan Gulzari[a*], Angelo Mencarelli[b*], Mark Camara[c], Chantal Roozeboom[a], Hans Komen[a], John W.M. Bastiaansen[a]

[a]Department of Animal Breeding and Genomics, Wageningen University and Research, Droevendaalsesteeg 1, 6708 PB, Wageningen (The Netherlands)

[b]Wageningen Plant Research, Wageningen University and Research, Droevendaalsesteeg 1, 6708 PB, Wageningen (The Netherlands)

[c]Wageningen Livestock Research, Wageningen University and Research, Droevendaalsesteeg 1, 6708 PB, Wageningen (The Netherlands)

*Shared first authorship

Submitted to Aquaculture

Abstract

Phenotyping can be a financial and logistical bottleneck for fish breeding programs. The most desired traits in breeding programs are difficult-to-measure since they require slaughtering fish or are time consuming, expensive, and laborious. Indicator traits measured on the selection candidates themselves can be used to genetically improve difficult-to-measure traits. Morphometric measurements can be used as indicator traits and computer vision simplifies and accelerates extracting many morphometric measurements simultaneously from digital images. The objectives of this study were to develop an automated process for the extraction of morphometric measurements from 2D and 3D images, to estimate their genetic parameters and genetic correlations with the production traits of interest, and to validate their quality as indicator traits for harvest and slaughter traits. We extracted 40 morphometric measurements from 2D and 3D images of 933 gilthead seabream including various distance, area, and volume measurements. We performed genetic analyses using a genomic relationship matrix. Volume measurements from 3D images were superior to 2D measurements as indicator traits for harvest weight and residual fillet weight. However, 2D measurements provided better indicators for fillet fat percentage. Morphometric traits had a wide range of heritabilities. *Tail excluded length* had the highest heritability (0.48 ± 0.05), and *Tail area* (0.06 ± 0.05) and *4th equidistant thickness* (0.04 ± 0.03) had the lowest heritabilities. The 3D morphometric traits *total volume* had a genetic correlation of 1.00 ± 0.002 with harvest weight and *trunk volume* had a genetic correlation of 0.89 ± 0.10 with residual fillet weight. In contrast, *fillet area,* a 2D trait had the strongest genetic correlation of -0.75 ± 0.15 with fillet fat percentage. Based on these strong genetic correlations, there is considerable value for using 2D and 3D image analysis to extract morphometric traits as indicator traits for harvest weight, residual fillet weight, and fillet fat percentage.

2.1 Introduction

Phenotyping can be a financial and logistical bottleneck for fish breeding programs. Except for growth rate, most desired traits in breeding programs are difficult-to-measure because they either require slaughtering fish or taking the measurements is time consuming, expensive, and laborious (Rodde et al., 2020). Typical examples of such difficult-to-measure traits include processing yield, product quality, and feed conversion efficiency (Chavanne et al., 2016). These traits are often not (yet) included in aquaculture breeding programs or are improved using sib data. Selection on sib data may limit genetic progress because only the between family variance is exploited or requires genomic selection to capture the within family Mendelian sampling variance. Alternatively, indicator traits measured on the selection candidates themselves capture the Mendelian sampling and can be used to genetically improve difficult-to-measure traits. However, the efficiency of selection using indicator traits depends on both their heritability and their genetic correlation with difficult-to-measure traits (Falconer and Mackay, 1996).

Ease of measurement is an important reason for using morphometric measurements as indicator traits. Computer vision simplifies and accelerates extracting many morphometric measurements simultaneously via image capture and subsequent extraction of information from digital images (Dowlati et al., 2012; Navarro et al., 2016). Computer vision has been used to extract morphometric measurements from two-dimensional (2D) images of common sole (*Solea solea*) (Blonk et al., 2010), common carp (*Cyprinus carpio*) (Prchal et al., 2018), Nile tilapia (*Oreochromis niloticus*) (Cardoso et al., 2021; Fernandes et al., 2020), rainbow trout (*Oncorhynchus mykiss*) (Haffray et al., 2013), and gilthead seabream (*Sparus aurata*) (Navarro et al., 2016). Ultrasound imagery has been used to extract internal measurements in European seabass (*Dicentrarchus labrax*) (Vandeputte et al., 2017) and rainbow trout (Haffray et al., 2013). Morphometric measurements from three-dimensional (3D) images of fish

have been performed in Liao et al., (2021); however, their value as indicator traits was not established.

For gilthead seabream, studies have been limited to morphometric traits from 2D images as predictors for harvest weight and fillet yield (Navarro et al., 2016). In this study we use 2D and 3D images of gilthead seabream to discover indicator traits for harvest weight, fillet yield, fillet fat percentage, and viscerosomatic index. The objectives of this study were to develop an automated process for the extraction of morphometric measurements from 2D and 3D images, to estimate their genetic parameters and genetic correlations with the traits of interest for aquaculture production, and to validate their quality as indicator traits.

2.2 Materials & Methods

2.2.1 Production sites and production of experimental fish

For this experiment, a population of juveniles was produced by natural mass spawning 33 males of 3 years old and 20 females of 4 years old. Eggs were collected as a single batch on one day. When fish reached an average weight of 3 grams, a batch of 84,650 juveniles was stocked in a sea cage in southeast of Spain and grown under commercial conditions.

2.2.2 Phenotypic data collection

We collected data on production traits from the commercially produced fish at harvest after a grow-out period of 500 - 504 days. In total, we sampled 945 fish for data collection over 5 consecutive days in September (daily 150 – 200 fish). We harvested a random sample of fish every morning from the cage into an oxygenated tank. Before processing, small batches of fish received a lethal dose of clove oil (0.03 mL/L). We repeated this process until all fish were measured for all the traits.

We measured body weight with a scale sensitive to 0.5 grams. Fillet fat estimations were taken on whole fish from eight points (four on each side of the fish) using a Distell Fish Fat meter (Distell Inc., West Lothian, Scotland) and averaging the eight measurements on the same fish. The fish were then gutted,

and viscera weight was recorded with a scale sensitive to 0.5 grams. Viscera included all the internal organs and abdominal fat. We calculated viscerosomatic index as (viscera weight / harvest weight) x 100. The gutted fish were then filleted, and fillet weight was recorded for the left side. The recorded fillet weight was multiplied by two to estimate the total fillet weight. Measurements taken on a single day were performed by the same person for each trait. Fillet yield was calculated as (total fillet weight/harvest weight) x 100.

We used R software and the "tidyverse" package collection to edit and organize the data (R Core Team, 2020; Wickham et al., 2019).

2.2.3 Image acquisition and morphometric traits

When the fish were whole, we acquired their images using an Intel® RealSense™ D435 Depth Camera (Intel Corp., Santa Clara, California, USA) fitted in a cabinet 50 centimeters from the bottom surface. The bottom surface of the cabinet was made of blue anti-reflective plastic to assist in the segmentation of the fish, and the walls and ceiling were made of white plastic to enhance light diffusion. The light source inside the box was white light producing LED stripes (OSRAM GmbH, Munich, Germany) covered with light diffusers. The resolution of the images was 1280 x 720 pixels for both 2D and 3D images. Examples of captured images can be found in Figure 2.1.

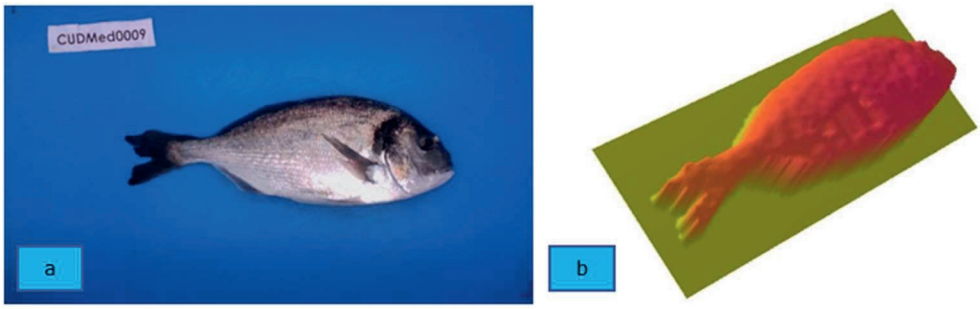

Figure 2.1. a) 2D color image of fish on a blue background with ID label. b) 3D image of fish.

From the 2D and 3D digital images, we automated the extraction of morphometric measurements with Halcon 20.11 framework (MVTec Software GmbH, Munich, Germany) (Eckstein and Steger, 1999). We chose the morphometric measurements that we expected to be biologically related to production traits, including some traits defined by Cardoso et al. (2021); Fernandes et al. (2020); Navarro et al. (2016); and Prchal et al. (2020). The process of extracting information from the digital images consisted of three sequential subprocesses: 2D and 3D image segmentation, landmark detection, and extraction of morphometric measurements (Figure 2.2).

2.2.3.1 Image segmentation

The RGB (2D) and depth (3D) image segmentation included four main sub-processes as depicted in Figure 2.2: RGB color space conversion, image segmentation and region orientation, RGB and depth image rotation, and RGB and depth image foreground and background separation. We converted the RGB image (Figure 2.1.a) into the Hue, Saturation, Intensity (HSI) color space (Gonzales and Woods, 2018) and automatically segmented the saturation channel and generated the binary fish region (FR) (see Appendix 2.6.2.1 for details). We calculated the orientation of the main axis of the FR to generate the affine transformation matrix for rotating the FR, the RGB image, and the depth image to generate FR_{aff}, RGB_{aff}, and $Depth_{aff}$, respectively (see Appendix 2.6.2.2 for details). After the image rotation, we used FR_{aff} as a mask to separate the foreground from the background in the RGB_{aff} and $Depth_{aff}$ images to generate RGB-FR_{aff} (Figure 2.3) and Depth-FR_{aff} (Figure 2.A.2).

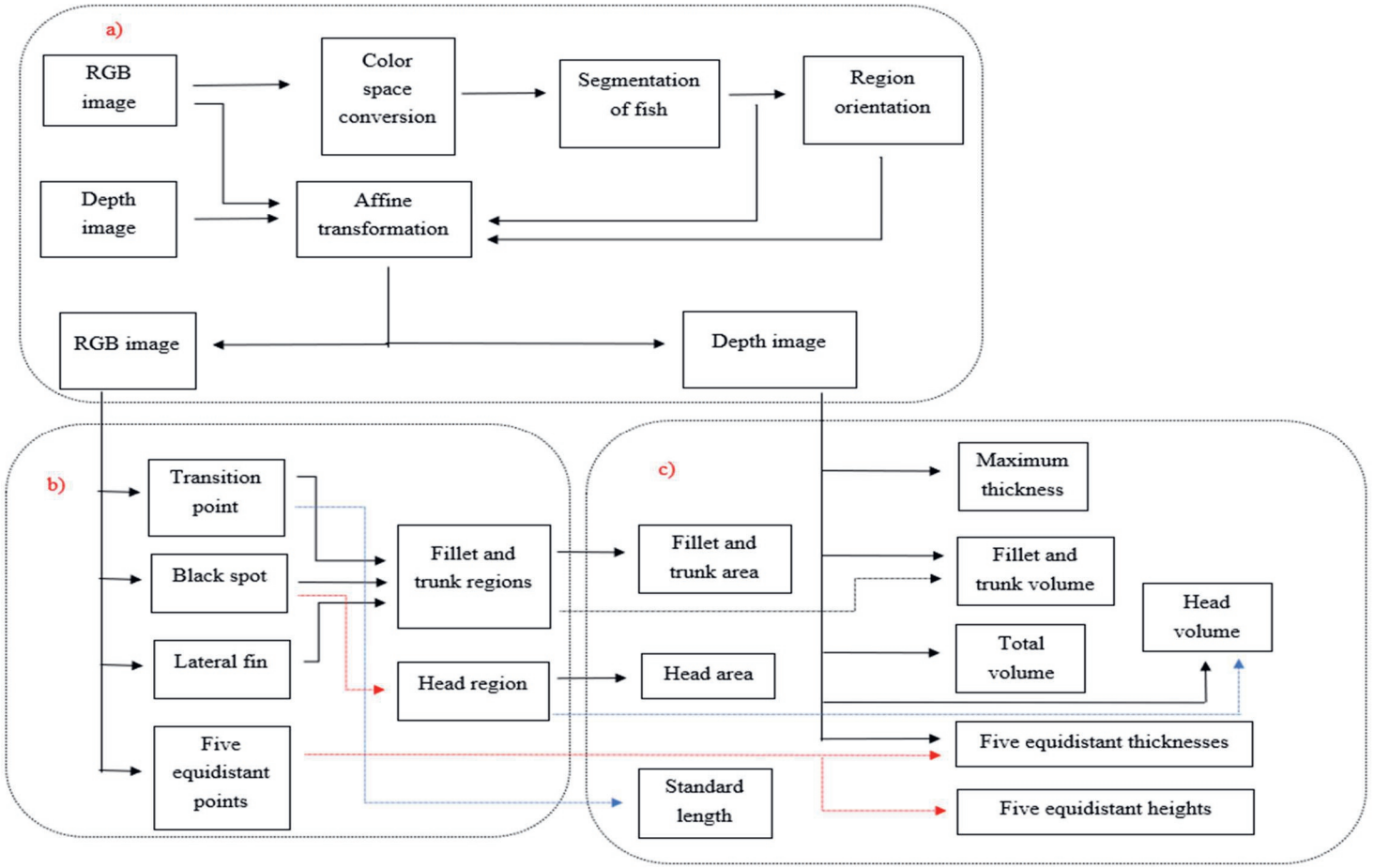

Figure 2.2. Three sequential subprocesses of automated image processing: a) RGB and depth image segmentation, b) landmark detection, c) extraction of morphometric measurements.

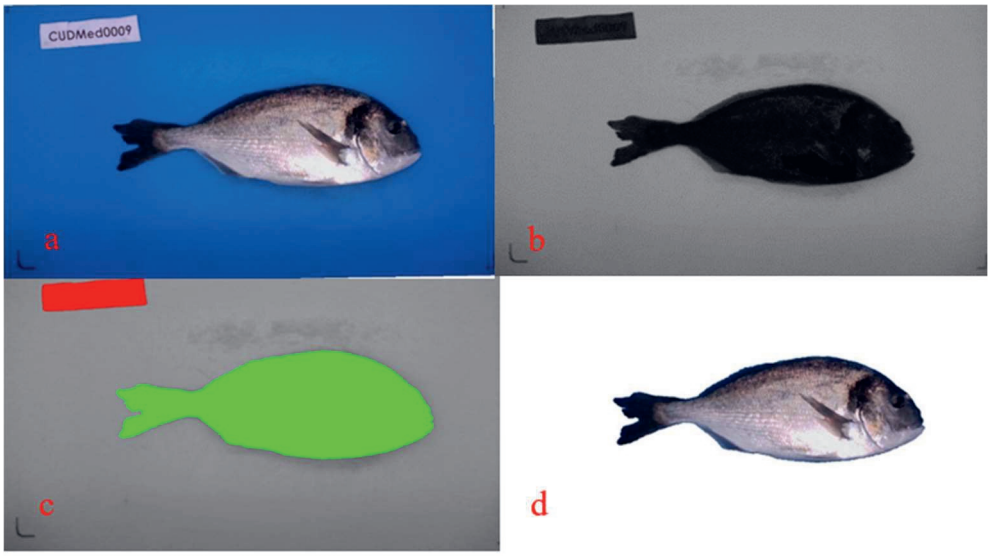

Figure 2.3. a) an RGB image. b) the saturation channel after color space conversion. c) after the segmentation: the ID label region (in red) and binary fish region (in green). d) the fish separated from background (RGB_FR_{aff}).

2.2.3.2 Landmark detection

In this paper, the term "landmark" is used to refer to specific points of interest in images that have clear anatomical features or locations. The term "pseudo-landmark" is used to refer to points that can only be defined by relative locations. We identified a set of landmarks and pseudo-landmarks in a sequential process, first in the RGB-FR_{aff} for the 2D landmarks and pseudo-landmarks, and subsequently we projected them on the Depth-FR_{aff} for the 3D feature extraction. Initially, we defined the transition point where the peduncle makes a transition to the caudal fin. We computed the FR_{aff} height by calculating the distance between each point of the FR_{aff} and its perpendicular projection onto the FR_{aff} main axis (Figure 2.4). The minimum value of FR_{aff} height gave the peduncle height (Appendix 2.6.3.1). We divided the region between the maximum height line and the tip of the head into four quadrants (Figure 2.A.3). The segmentation of blue-channel-enhanced image in the first quadrant individuated the black spot

(Appendix 2.6.3.2). The black spot is marked in yellow in Figure 2.A.3.b. The purpose of identifying the black spot and its quadrants was to have a

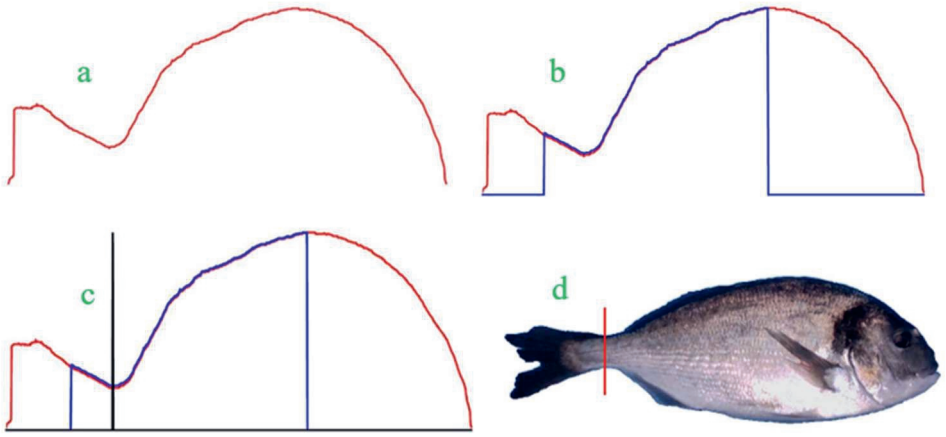

Figure 2.4. a) plot of the $Depth_{aff}$ thickness. b) the clipped region of the FR_{aff} thickness distance. c) computed transition point defined by the minimum value of the FR_{aff} thickness. d) its correspondent vertical line passing through the coordinate of the transition point.

reproducible landmark to locate several morphometric measurements. The centroid of the black spot defined the vertical line that divides the head region from the rest of the body (Figure 2.A.3). We defined the trunk region as the body region between the transition point and the head region. We computed the entropy of the gray values of the blue-channel-enhanced RGB_{aff} by using a 17 x 17 pixels kernel (Sparavigna, 2019). Then we reduced the resulting image domain to the fourth quadrant of the frontal region (Figure 2.A.4.a). After automatic thresholding, the lateral fin region was determined (Figure 2.A.4.b). Finally, the bottom left coordinate of the smallest rectangle orthogonal to the image reference system (IRS) around the lateral fin region defined the passing point of the horizontal line of the border between the two regions, the ventral region (Figure 2.A.4.c) and the fillet region (Figure 2.A.4.d). The five equidistant points

(FEP) (Navarro et al., 2016) were automatically generated on the FR_{aff} main axis and the coordinates of the intersections of the vertical lines passing through the FEP and the contours of the FR_{aff} were computed (Figure 2.A.5). We computed the FEP heights on the FR_{aff} ($FEP_{2D\text{-}heights}$) from the FEP coordinates. We computed the depth along each line by projecting the five vertical lines on the Depth-FR_{aff} and defined the maximum depth as the thickness of the FEP expressed in millimeters ($FEP_{3D\text{-}thickness}$) (Figure 2.A.5.b).

2.2.3.3 Extraction of morphometric measurements

We combined the landmarks and pseudo-landmarks to extract three types of morphometric measurements: 1) distances as linear measures (in millimeters as one-dimensional measures); 2) areas as quadratic measures (in millimeters squares as two-dimensional measures); 3) volumes as cubic measures (in cubic centimeters as three-dimensional measures).

Detailed explanations of all the steps of extracting morphometric measurements can be found in Appendix 6.4. In total, 40 morphometric measurements were extracted. Morphometric measurements on the RGB image can be found in Figure 2.5 and morphometric measurements on the depth image can be found in Figure 2.6.

From those measurements, we calculated *relative head area* as *head area / total area*, *relative fillet area* as *fillet area / total area, thickness by height* as *thickness / height*, and *height by length* as *height / length*. We calculated *fish* and *head eccentricities* as $e = \sqrt[2]{1 - \left(\frac{a}{b}\right)^2}$, where *a* and *b* are the main and the secondary radius of the ellipse having the same orientation and aspect ratio as the *fish area* and the *head area*, respectively as defined by Fernandes et al., (2020). The *trunk area* represents *fillet area + ventral area.* From 3D images, we measured *relative ventral volume* as *ventral volume / total volume,* and *ventral volume by fillet volume* as *ventral volume / fillet volume*. We measured the *equidistant thicknesses* as the distance between the left and right side of the fish at the FEP

(Figure 2.6.d). The calculation of volume was adjusted for the roundness of the fish as depicted in Figure 2.7.

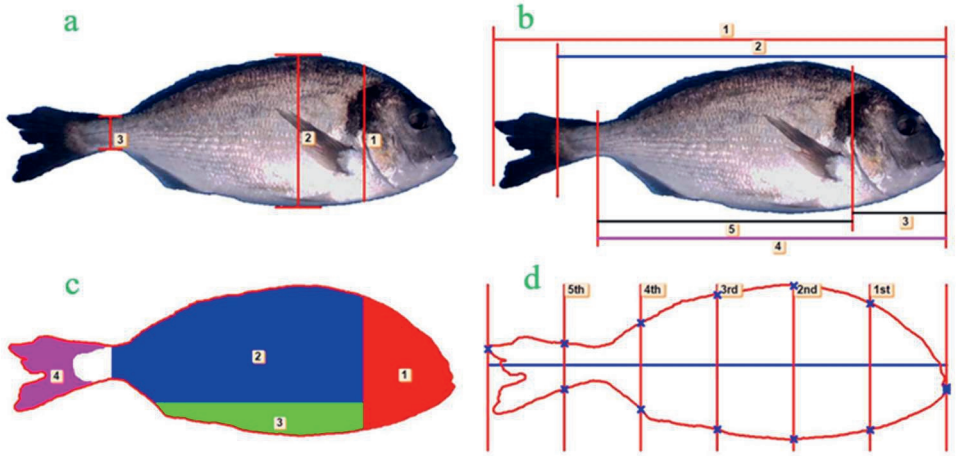

Figure 2.5. 2D morphometric measurements. a) 1 = head height, 2 = maximum height of fish, 3 = peduncle height,

b) 1 = total length, 2 = standard length, 3 = head length, 4 = tail excluded length, 5 = fillet length,

c) 1 = head area, 2 = fillet area, 3 = ventral area, 4 = tail area,

d) between the blue points of each line are equidistant height measurements.

2.2.4 Genetic analyses

2.2.4.1 DNA extraction and genomic relationship matrix

IdentiGEN (Dublin, Ireland) isolated the DNA from fin clips and performed the genotyping using the ~30k "MedFish" SNP array (Peñaloza et al., 2021). We filtered the genotypic data by excluding SNPs that had call rates below 90%, were not segregating, or had Hardy-Weinberg equilibrium exact test p-values below 1e-10.

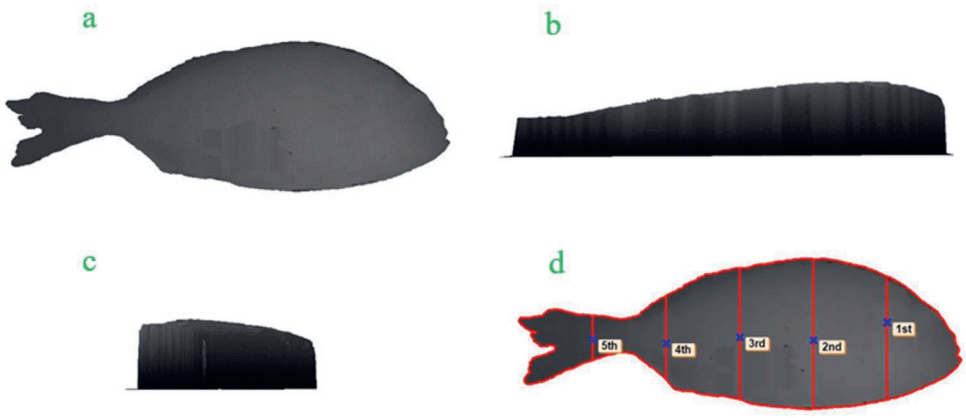

Figure 2.6. 3D morphometric measurements: a) top projection of the foreground of the depth image, b) the lateral view, c) the front view, d) the FEP thicknesses computed on the foreground of the depth image.

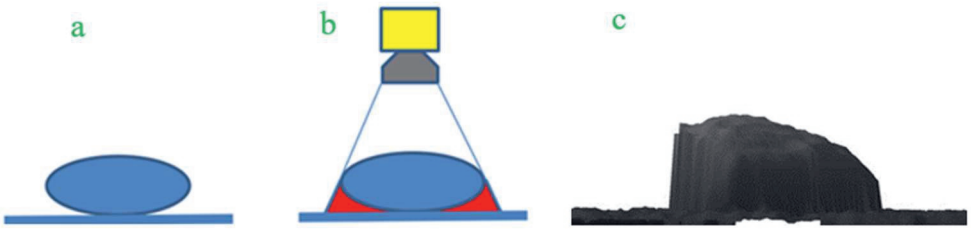

Figure 2.7. a) an example of an object with an oval frontal section. b) how the camera captures only one side of the object and projection of that side on the background. c) the total computed volume was the composition of the volume of the object (blue section area) and the projected volume (red section area).

We computed a genomic relationship matrix (GRM) based on the remaining 28164 SNPs using calc_grm software (Calus and Vandenplas, 2019) with the "vanraden" option (VanRaden, 2008), in which the GRM is computed as $G = \frac{ZZ'}{2\sum_{p_i} p_i (1-p_i)}$. To use this option, we coded marker genotypes as "0", "1", or "2". Z is

a matrix that contains marker genotypes for all loci, which is corrected for the allele frequency per locus. p_i is the frequency of the minor allele. The inverse of the GRM matrix was obtained directly from calc_grm using "giv" function.

2.2.4.2 Estimation of genetic parameters

We estimated the heritabilities of the morphometric measurements using univariate models, and genetic (r_g) and phenotypic (r_p) correlations between phenotypes and morphometric measurements using bivariate animal models with ASReml version 4.2 (Gilmour et al., 2015).

The univariate animal model is $\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{Zu} + \boldsymbol{e}$, where y is a vector of a morphometric measurement or production trait, β is the vector of fixed effects sampling day and harvest weight as a covariate, u is the vector of random animal additive genetic effects $\sim(0, G\sigma_a^2)$, where G is the GRM and σ_a^2 is the additive genetic variance of a morphometric trait, and e is the vector of random residual effects $\sim(0, I\,\sigma_e^2)$, where I is an identity matrix and σ_e^2 is the residual variance of a morphometric trait. X and Z are design matrices, that relate observations to the fixed effect and additive genetic effect of animals, respectively. We included harvest weight as a covariate to obtain the genetic variance of morphometric traits and production traits that are independent of harvest weight. Naturally, the genetic model of harvest weight included "sampling day" as the only fixed effect.

The bivariate animal model is $\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{Zu} + \boldsymbol{e}$, where y is a vector of the phenotypes of a production trait and a morphometric measurement, β is either the vector of fixed effect "sampling day" in the models with harvest weight as the production trait or the vector of fixed effect sampling day and covariate harvest weight in the models with fillet weight, viscera weight, and fillet fat percentage as the production trait, u is the vector of random animal additive genetic effects $\sim\left(\begin{bmatrix}0\\0\end{bmatrix}, G\begin{bmatrix}\sigma_{a,T_P}^2 & r_{a,T_{P,M}}\sigma_{a,T_P}\sigma_{a,T_M}\\ r_{a,T_{P,M}}\sigma_{a,T_P}\sigma_{a,T_M} & \sigma_{a,T_M}^2\end{bmatrix}\right)$, where G is the GRM and σ_{a,T_P}^2 is the additive genetic variance of a production trait (e.g. harvest weight), σ_{a,T_M}^2 is the additive genetic variance of a morphometric trait (e.g. total

volume), $r_{a,T_{P,M}}$ is the additive genetic correlation between production trait and morphometric trait, and e is the vector of random residual effects $\sim\left(\begin{bmatrix}0\\0\end{bmatrix}, I\begin{bmatrix}\sigma^2_{e,T_P} & r_{e,T_{P,M}}\sigma_{e,T_P}\sigma_{e,T_M}\\ r_{e,T_{P,M}}\sigma_{e,T_P}\sigma_{e,T_M} & \sigma^2_{e,T_M}\end{bmatrix}\right)$, where I is an identity matrix, σ^2_{e,T_P} is the residual variance of a production trait, σ^2_{e,T_M} is the residual variance of a morphometric trait, and $r_{e,T_{P,M}}$ is the residual correlation between the production trait and the morphometric trait. X and Z are design matrices, that relate observations to the fixed effect and additive genetic effect of animals, respectively. With harvest weight as a covariate in the models for fillet weight, the trait becomes residual fillet weight, which is considered as a proxy for fillet yield. Likewise, residual viscera weight becomes a proxy for viscera yield. In all models, we included "sampling day" as a categorical variable with one category for each measurement day (5 days), which includes the effect of the person who performed the phenotyping of production traits because measurements taken on a single day were performed by the same person for each trait. Gilthead seabream is a protandrous fish (Loukovitis et al., 2011) and all individuals were males during harvest. Therefore, we did not include "sex" as a fixed effect in this study. We excluded the *5th equidistant thickness* from the results because of convergence problems in genetic models that included this morphometric trait.

2.3 Results

2.3.1 Image analysis and traits

In total, 23 images were not used for the analyses, two due to the presence of the operator's hand, and 21 because they were images of the same fish taken two times. Each morphometric measurement was successfully extracted from all the selected images, although outliers appeared in some cases. Five outliers occurred due to the presence of an unexpected artifact during one of the acquisition sessions: in images where the fish was not placed in the center of the blue background, a small black square of 5 x 5 mm was visible and was touching the dorsal side of the fish. The black square was detected as part of the fish and

caused overestimation of some measurements. In some other cases, the doors of the imaging cabinet were not shut properly, and the outside light illuminated the image, causing differences in color and brightness. For those images, we adjusted the contrast in the automated image analysis process. Values for each morphometric measurement that were more than 3.5 standard deviations from the mean were re-coded as missing values. The image acquisition time was up to 10 seconds to put the fish on the device, take the image, and remove the fish. The average time of image processing was 344.2 (± 23.9) milliseconds, performed in a laptop equipped with eight processors Intel® Xeon® CPU E3-1505M v5, 2.80 GHz. The choice of Intel® RealSense™ D435 depth camera was based on its compactness, minimal distance required from the depth camera to the scene (~280 mm with a resolution of 1280×720), and favorable price-quality ratio. MVTec Halcon software was chosen because it is an industrial software for machine vision specifically suited for the development of image processing solutions, image segmentation and metrology from images. The coefficient of variation of the 24 distance traits had a median of 5.33 and ranged from 4.06 to 10.5 with two exceptions for *fish eccentricity* (CV = 0.80) and *4th equidistant thickness* (CV = 22.8). CV of the nine area traits were higher with a median of 11.2 (5.96 – 23.1), and for the seven volume traits even higher with a median of 17.7 (17.1 – 27.4). The number of records available for each morphometric measurement and summary statistics can be found in Supplementary Table 2.A.1. Pearson correlations between morphometric measurements can be found in Supplementary Table 2.A.2.

2.3.2 Genetic parameters

The heritabilities of morphometric traits ranged from low to high (Supplementary Table 2.A.3). *Tail excluded length* had the highest heritability (0.48 ± 0.05). *Tail area* (0.06 ± 0.05) and *4th equidistant thickness* (0.04 ± 0.03) had the lowest heritabilities.

Several morphometric traits were strongly genetically correlated to production traits (Table 2.A.4). The 3D morphometric trait, *total volume* had a perfect genetic correlation with harvest weight (1.00 ± 0.002) (Table 2.1), showing that they are genetically the same trait. Another 3D morphometric trait, *trunk volume,* had the highest genetic correlation to residual fillet weight (0.89 ± 0.10) (Table 2.1) where *relative fillet area*, was the strongest correlated 2D trait but with a weaker genetic correlation (-0.77 ± 0.12). The *fillet area* (2D) had the strongest genetic correlation with fillet fat percentage (-0.75 ± 0.15), and *3rd equidistant thickness* (1D) had slightly lower genetic correlation (0.72 ± 0.09). The *tail excluded area* had the strongest genetic correlation (-0.37 ± 0.12) with residual viscera weight (among the genetic correlations that were significantly different from zero). All genetic and phenotypic correlations between morphometric traits and production traits are reported in Supplementary Table 2.A.4 and 2.A.5, respectively.

Table 2.1. Distance, area, and volume morphometric traits with the strongest genetic correlations with production traits.

Production trait	**Morphometric trait***	r_g#
Harvest weight	Total volume	1.00 ± 0.002
	Trunk area	0.98 ± 0.01
	Maximum height	0.97 ± 0.01
Residual fillet weight	Trunk volume	0.89 ± 0.10
	Relative fillet area	-0.77 ± 0.12
	Head length	-0.68 ± 0.13
Residual fillet fat percentage	Fillet area	-0.75 ± 0.15
	3rd equidistant thickness	0.72 ± 0.09
	Head volume	-0.47 ± 0.12
Residual viscera weight	Tail excluded area	-0.37 ± 0.12
	Total volume	-0.37 ± 0.20
	Perimeter length	-0.34 ± 0.13

*All distance and area morphometric measurements were in millimeters and squared millimeters, respectively, volume measurements were in cubic centimeters; r_g = Genetic correlations of morphometric traits to production traits

#Harvest weight was a covariate in the genetic models for all production traits except for harvest weight

The genetic parameters of production traits are presented in Table 2.2. The production traits had high heritabilities except for residual fillet weight.

Table 2.2. Genetic and environmental variance, and heritability estimates of production traits.

Production trait*	V_A	V_E	h^2 ± se
Harvest weight	2796	2417	0.54 ± 0.05
Residual fillet weight	14.2	118.6	0.11 ± 0.04
Residual viscera weight	9.1	8.8	0.51 ± 0.05
Residual fillet fat percentage	3.4	2.7	0.56 ± 0.05

*Harvest weight was a covariate in the genetic models for all production traits except for harvest weight

2.4 Discussion

In this study, we introduced a fast, objective, and reproducible image analysis method that extracts morphometric measurements simultaneously from 2D and 3D images of seabream. The extraction process is fully automated in the sense that after a fish image is available, the morphometric measurements are extracted using fully automated image analysis. We tailored the image analysis methods to images of gilthead seabream. For instance, the identification of specific landmarks such as the black spot is specific to gilthead seabream and not applicable to most other species. However, the presented methods can be modified easily to identify specific landmarks in the species of interest. If desired, additional morphometric measurements can be extracted for genetic analyses.

Phenotyping with image analysis has several advantages. Automated phenotyping provides more accurate measurements than manual phenotyping because the measurements do not depend on the person who performs the phenotyping (Hao et al., 2016) whereas manual measurements are prone to human error. For example, Rutten et al. (2004) reported that the person who fillets the fish has considerable influence on fillet yield. In our data, the effect of person was not estimated separately but confounded with measurement day, which had a strong significant effect on residual fillet weight. Reduced measurement errors are expected to yield higher heritabilities. Elalfy et al. (2021) reported that automatic measurements of morphometric traits yielded on average 24% higher heritabilities than manual measurements. In addition, León-Bernabeu et al. (2021) reported that *fork length* had 88% higher heritability when automatically measured. Another advantage of image analysis is that performing external measurements on large numbers of fish is fast and simple. In addition, the images of fish can be stored and used to obtain measurements in the future because even if the fish is no longer available, the images can be analyzed later to create ancestral data for genetic analyses. Besides phenotyping for genetic evaluation in breeding programs, the external measurements can also be used in different settings such as processing or to quantify the shape and value of the fish for marketing.

Genetic analyses of image-derived morphometric measurements have been performed to discover high quality morphometric indicator traits for breeding goal traits (Cardoso et al., 2021; Haffray et al., 2013; León-Bernabeu et al., 2021; Prchal et al., 2020, 2018). The heritability of some morphometric traits reported here are lower than the values reported in the literature. In this study, we estimated the heritabilities of morphometric traits with harvest weight included as a covariate in the genetic models. These models yield genetic variation of morphometric traits after variation due to harvest weight is accounted for. Harvest weight was included as a covariate to obtain heritability estimates for

morphometric traits that validate their quality as indicator traits. Therefore, we see a relatively low heritability of *total area* (0.37 ± 0.05), lower than the heritability estimated by León-Bernabeu et al. (2021) (0.51 ± 0.10), based on automated measurements on gilthead seabream. The heritabilities of *equidistant height* traits were also lower (0.27 - 0.33) than the heritabilities estimated by León-Bernabeu et al. (2021) (0.34 - 0.59). Most likely, these lower heritabilities are because of our inclusion of harvest weight as a covariate in the genetic models unlike in León-Bernabeu et al. (2021). In this study, *tail area* had the lowest heritability among all area measurements (0.06 ± 0.05), which is much lower than the heritability of 0.21 estimated by Cardoso et al. (2021) in Nile tilapia. The low heritability of tail area might be caused by environmental factors that obscure genetic variance such as tail damage. In this study, *tail excluded length* had the highest heritability among all morphometric traits, which is not surprising because it is a purely skeletal measure, which typically have the highest heritabilities (Bourdon, 2000). The heritability of *total length* was lower than *tail excluded length*, which may also be due to variation caused by tail damage. The heritabilities of *total volume* (0.07 ± 0.03) and *trunk volume* (0.12 ± 0.04) were very low, but not surprising given their almost unity genetic correlations to harvest weight.

The quality of morphometric indicator traits, as measured by their genetic correlations with the production traits, was found to be very high for harvest weight, high for residual fillet weight and residual fat percentage, and much lower for viscera weight. *Total volume* is a perfect indicator for harvest weight. Although weighing a fish is simple, considerable effort is necessary if there are thousands of fish in a breeding program. When fish are imaged and phenotyped in an automated way, taking a manual weight is no longer needed.

In contrast to harvest weight the indicator traits for residual viscera weight have relatively low genetic correlations. One reason is that harvest weight is included as a covariate, but we still see the strongest genetic correlation with *tail excluded*

area and *total volume* (-0.37). The correlation with *total volume* is interesting because it has a high correlation with the covariate harvest weight. It is possible that the part of the variation in volume that is independent of body weight is particularly correlated to viscera weight. The density of viscera may vary more than the density of other tissues, due to for instance variable presence of feed and water in the gut, and variable size of fat deposits relative to other organ tissues.

Fillet area has the highest genetic correlation to residual fillet fat percentage (-0.75 ± 0.15). The *fillet area* overlaps with the dorsal part of the fillet, which means the fat content of the fillet will reduce if the fillet is larger at the dorsal part. On the contrary, the fat content of the fillet will increase if the fillet is larger at the ventral part as *ventral area* is positively correlated to fillet fat percentage. Among the distance traits, the *3rd equidistant thickness* has the highest genetic correlation to fillet fat percentage, which means that a thicker fish is expected to contain more fat in the fillet. It is easy to see that more fat could lead to a thicker fish. Possibly this increased *thickness* affects the position of the horizontal line that separates the *fillet area* and the *ventral area*, making those areas valuable indicator traits for residual fillet fat percentage.

Residual fillet weight, our proxy for fillet yield, is arguably the most important production trait we analyzed because fillet is the most valuable part of the fish. We found a strong genetic correlation of residual fillet weight with *trunk volume* (0.89 ± 0.10). Correlations of volume traits with fillet yield have not been reported previously but we can compare with measures of thickness and area. The *3rd equidistant thickness* is positively correlated to residual fillet weight (0.61 ± 0.17) in agreement with Rutten et al. (2005), who reported a strong positive genetic correlation (0.98) between thickness and fillet yield in Nile tilapia. Harvest weight was not a covariate in the genetic models of Rutten et al. (2005), which may explain in part the lower correlation we found.

Among the length traits, *head length* had the highest genetic correlation to residual fillet weight (-0.68 ± 0.13). Morphometric traits of the head have been reported to be correlated to fillet yield in several species including common carp (Prchal et al., 2018), sea bass (Vandeputte et al., 2017), and rainbow trout (Haffray et al., 2012). Others found no genetic correlation between head morphometric traits and fillet yield in rainbow trout (Kause et al., 2007) and Nile tilapia (Rutten et al., 2005). Species specific shapes and body proportions of different fish may lead to different genetic correlations of morphometric traits with production traits. Hence, genetic correlations with morphometric traits should be validated separately for each fish species.

In this study, *relative fillet area* was negatively correlated to residual fillet weight (-0.77 ± 0.12). This means that if the *relative fillet area* is larger, residual fillet weight is expected to be lower, which is counter intuitive. In other words, when *relative fillet area* goes up, harvest weight increases faster than fillet weight. We also found that the genetic correlation of *ventral volume by fillet volume* with residual fillet weight was positive, meaning that if the ventral volume increases compared to fillet volume, residual fillet weight is expected to be higher. These results could mean that most of the genetic variation in residual fillet weight is in the ventral part of the fillet. We did not weigh the top and bottom part of the fillet separately, but we do see that the genetic variation of *ventral volume* was about twice that of *fillet volume* while the phenotypic variance of *fillet volume* is about twice that of *ventral volume* (Table 2.A.3). Previously we showed that viscera weight has low variation (Gulzari et al., 2022), therefore the high genetic variation of *ventral volume* is probably caused by variation in the fillet tissue. In this case, then the response to selection for residual fillet weight may be seen mostly in the ventral part of the fillet.

We find that volume measurements from 3D images emerge as superior indicator traits for harvest weight and residual fillet weight compared to 2D measurements (Table 2.1). A novelty of this study is that we extracted the volume

measurements directly from the 3D images as opposed to estimating them by combining 2D morphometric measurements as performed in Navarro et al., (2016). Although we cannot compare the precision of volume measurements with other studies, we expect that direct measurement of volumes from 3D images is more precise than estimating them from 2D images. Direct measurement of volumes involves no extra effort because all morphometric measurements are extracted simultaneously, but it does require a 3D image, and therefore an initial investment for the 3D imaging equipment.

Image analysis is preferable to manual phenotyping in gilthead seabream breeding programs, even for non-invasive traits such as body weight. Images are easy to take with the right equipment, and image analysis offers precise, fast, and uncomplicated extraction of morphometric measurements. Several morphometric measurements can be used as high-quality indicator traits for production traits such harvest weight, fillet yield, and fillet fat percentage (Table 2.1). For invasive traits, image analysis has the additional advantage of performing measurements on selection candidates rather than having to rely on sib data.

We used classical segmentation and metrological approaches to extract information from 2D and 3D images, which allowed comparison to previous studies and interpretation of the indicator traits. In recent literature, machine vision is combined with deep learning (DL) models to predict target traits directly from images. While this works for 2D images, DL models currently do not cope with 3D image measurements. Xue et al. (2023) built a DL model that learns from raw images without explicit annotation of landmarks, and they demonstrated that DL could predict phenotypes of several traits without direct measurements of morphometric traits. Advantages of DL models include the robustness to variations in image quality such as differences in illumination and contrast (Akhtar and Mian, 2018), and the presence of occlusion or overlapping between fishes (van Essen et al., 2021). In our study we developed a well-defined and

controlled environment with diffuse and stable light, and calibrated dimensions that work well with classical segmentation. To segment parts of the fish such as the black spot and the lateral fin, a semantic segmentation approach can be used (Petrellis, 2021). In some cases, such as an industrial deployment where image analysis must cope with different or mixed species per session of measurements, a DL approach will be essential. However, there are no real advantages of training a neural network for the features that could be segmented in a classical way.

2.5 Conclusions

Automated phenotyping with 2D and 3D images of gilthead seabream can extract many morphometric measurements in less than a second. Volume measurements from 3D images offer high quality indicator traits for harvest weight and residual fillet weight. In addition, we validated the value of several 2D measurements as indicator traits for residual fillet fat percentage in sea bream. Image analysis is preferable to manual phenotyping in gilthead seabream breeding programs, even for non-invasive traits such as body weight.

2.6 Appendix

2.6.1 Image calibration

Extraction of information from digital images started with the image calibration process. The RGB (2D) and the Depth (3D) images were directly aligned during the acquisition by using the Application Programmable Interface (API) functions of the Intel® RealSense™ Library (Intel Corp., Santa Clara, California, USA). Although the RealSense camera was delivered already calibrated and the stereo data can be given in millimeters, a calibration procedure was applied before beginning of image acquisition for sanity check and color calibration. For the calibration, an X-Rite ColorChecker® Classic was used that had 24 natural objects with chromatic, primary, and gray scale colors (Figure 2.A.1).

2.6.2 Image processing

Image segmentation and feature extraction from digital images were performed by using the Halcon 20.11 framework (MVTec Software GmbH, Munich, Germany) (Eckstein and Steger, 1999). The automated image processing consisted of three sequential sub-processes: RGB and depth image segmentation, landmark detection, and feature extraction (Figure 2.2).

2.6.2.1 RGB color space conversion and image segmentation

To perform image segmentation the RGB image was decomposed in red, green, and blue monochromatic channels, (R, G and B respectively) and then the RGB was converted into the Hue, Saturation, Intensity (HSI) color space (Gonzales and Woods, 2018) (Figure 2.3) with the transformation matrix

$$\begin{pmatrix} H \\ S \\ I \end{pmatrix} = \begin{pmatrix} \tan^{-1}(M2,M1) \\ \sqrt{M1^2+M2^2} \\ \frac{I1}{\sqrt{3}} \end{pmatrix} \text{ where } \begin{pmatrix} M1 \\ M2 \\ I1 \end{pmatrix} = \begin{pmatrix} \frac{2}{\sqrt{6}} & \frac{-1}{\sqrt{6}} & \frac{-1}{\sqrt{6}} \\ 0 & \frac{1}{\sqrt{2}} & \frac{-1}{\sqrt{2}} \\ \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} \end{pmatrix} \begin{pmatrix} R \\ G \\ B \end{pmatrix}$$

with range of values $H \in [0; 2\pi[, S \in \left[0; \sqrt{\frac{2}{3}}\right], I \in [0; 1]$.

After the conversion, the saturation channel was used to separate the fish from the background (Figure 2.3). The foreground, i.e., fish and the ID label, was segmented by using an automatic binary thresholding on the gray-level histogram according to Otsu algorithm (Otsu, 1979). The Otsu algorithm works by calculating the histogram of the image and then uses statistical moments to find the optimal threshold that divides the pixels into foreground and background and maximizes the separability between these two classes. From the heuristic knowledge that the ID label was always smaller than the fish, the fish region (FR) was selected in function of its shape and size (Figure 2.3).

2.6.2.2 Region orientation and image rotation

The cartesian coordinate system was used with origin on the top left of the image frame and with the x-axis and y-axis the axes of the system parallel to the horizontal and vertical axes of the image frame, respectively. The orientation of a region was defined as the angle between the main axis of a region and the x-axis. The orientation of the FR was calculated, an ellipse having the same orientation and the same aspect of the FR was used. From the FR orientation, the affine transformation matrix was calculated (Gonzales and Woods, 2018).

Finally, with the affine transformation matrix, the RGB and the Depth images were rotated around the center of the FR to create RGB_{aff} and $Depth_{aff}$ respectively. After the image rotation, foreground and background separation was performed as described in section 2.3.1.3.

2.6.3 Landmark and pseudo-landmark detection

After the foreground and background separation, landmarks and pseudo-landmarks were detected as described in section 2.3.2.

2.6.3.1 Transition point

To find the transition point, four steps were implemented. First, starting from the outer contour of the FR_{aff}, the smallest rectangle orthogonal to the image reference system (IRS) around the FR_{aff} is found (Figure 2.A.2.a). The rectangle

together with FRaff providethe coordinates of the extreme points of the FR_{aff} namely the tip of the tail (tip_t), the tip of the head (tip_h), top (top_f) and the bottom of the fish (bottom_f). Second, applying a minimum threshold of 2 mm to the Depth-FR_{aff}, a rough segmentation of fish body without caudal fin was performed and around it a second smallest rectangle orthogonal to the IRS is computed. From the coordinates of this rectangle the coordinates of the tip of the rough segmented fish body without caudal fin, tip-of-the-body-without-caudal-fin, is computed (Figure 2.A.2.b). Third, from the top-f and bottom-f x coordinates, the average x coordinate *top-bottom-average-x-coord* was computed. Subsequently, the *maximum height line* defined as the vertical line passing through the points with coordinates (*top-bottom-average-x-coord, top-f-y-coord*) and (*top-bottom-average-x-coord, bottom-f-y-coord*) was created, where *top-f-y-coord* and *bottom-f-y-coord* are the y coordinates of *top-f* and *bottom-f*, respectively (Figure 2.A.2.d). Fourth, the thickness of the FR_{aff} along its main axis for each pixel was computed. The thickness at one point on the main axis of a region is defined as the distance between the intersection of the contour with the plumb on the main axis in the respective point which are the furthest apart. Last, the FR_{aff} thickness was clipped between the x coordinates of the *tip-of-the-body-without-caudal-fin* and the x coordinate of the *maximum height line* (Figure 2.A.2.d). The clipped data presents only one local minimum (Figure 2.4.c), that is defined as transition point landmark (Figure 2.4.d).

2.6.3.2 Black spot, head region and trunk region

First, cropping the region of the FR_{aff} with the rectangle defined by the coordinates of the intersection of the maximum height line (Figure 2.A.3.a), with contour of the FR_{aff} and the coordinates of the *tip-h*, the frontal region was defined.

Subsequently, the frontal region was divided in quadrants by the vertical and horizontal line passing through the geometrical center of the above rectangle

(Figure 2.A.3.a). Defining as initial quadrant the top left quadrant and moving clockwise, the first to the fourth quadrant are defined.

Second, the R, G, and B channels were combined to enhance the blue channel according to Excessive Green method (Meyer et al., 1998) in the following the formula:

$$EB = B \cdot B - R - G$$

where R, G, and B are the monochromatic images given by the R, G, and B channels of the RGB_{aff}, and EB is the Excessive Blue *i.e.,* enhanced blue channel. In EB the contrast between the black spot region and the surrounding tissues was enhanced. By reducing the blue-channel-enhanced domain to the first quadrant of the frontal region, more than 80% of the black spot region was inside the first quadrant (Figure 2.A.3.b). By automatically thresholding the first quadrant, the black spot region was segmented and, subsequently, its centroid computed (Figure 2.A.3.c). Third, cropping the *frontal region* with the rectangle defined by the coordinates of the intersection of the vertical line passing through the centroid of *black spot* region with contour of the *frontal region* and the coordinates of the *tip-h*, the *head region* was defined (Figure 2.A.3.d). Finally, cropping the FR_{aff} with rectangle defined by the coordinates of transition point and centroid of *black spot* region landmarks, x coordinates, and *top-f* and *bottom-f* pseudo-landmarks, y coordinates, the *trunk region* was defined (Figure 2. 5.c). Lateral fin, fillet and ventral regions were detected as described in section 2.3.2.3. Five equidistant points were detected as described in section 2.3.2.4.

2.6.4 Extraction of morphometric measurements

In this process, the landmarks and pseudo-landmarks are combined to compute the morphometric measurements. The main measurements are described in the following subparagraphs.

2.6.4.1 Distance (1D) features

All 1D features were given in millimeters. In Figure 2.5, the main vertical and horizontal features were shown.

Fish maximal height was given by the distance between the points with coordinates (*top-bottom-average-x-coord, top-f-y-coord*) and (*top-bottom-average-x-coord, bottom-f-y-coord*), where: *top-bottom-average-x-coord* was the average between the *top-f-x-coord* and *bottom-f-x-coord*, and *top-f-y-coord* and *bottom-f-y-coord* were the y coordinates of *top-f* point and *bottom-f* point, respectively (Figure 2.5.a).

The peduncle height (Figure 2.5.a) was the distance between the points with coordinates (*top-transition-point-x-coord, top-transition-point-y-coord*) and (*top-transition-point-x-coord, top-transition-point-y-coord*) where the coordinates were given by the intersection of the vertical line passing through the minimum value of the FR_{aff} thickness (Figure 2.A.2.c), and the contour of the FR_{aff} (Figure 2.A.3.b).

Total length (Figure 2.A.2.c) was the distance between the points with coordinates (*tip-tail-x-coord, tip-tail-and-tip-head-average-y-coord*) and (*tip-head-x-coord, tip-tail-and-tip-head-average-y-coord*), where: *tip-tail-and-tip-head-average-y-coord* was the average between the *tip-tail-y-coord* and *tip-head-y-coord*, and *tip-tail-x-coord* and *tip-head-x-coord* were the x coordinates of *tip-tail point* and *tip-head point* (Figure 2.A.2.c).

Standard length (Figure 2.A.2.c) was the distance between the points with coordinates (*tip-of-the body-without-caudal-fin-x-coord, tip-of-the body-without-caudal-fin-and-tip-head-average-y-coord*) and (*tip-head-x-coord, tip-of-the body-without-caudal-fin-and-tip-head-average-y-coord*), where *tip-of-the body-without-caudal-fin-and-tip-head-average-y-coord* was the average between the *tip-of-the body-without-caudal-fin-y-coord* and *tip-head-y-coord*, and *tip-of-the body-without-caudal-fin-x-coord* and *tip-head-x-coord* were the x

coordinates of *tip-of-the body-without-caudal-fin* point and *tip-head* point (Figure 2.A.1.b-c).

The Five Equidistant Points Heights (Figure 2.A.5.a) were given by distance between each $FEP_{2D\text{-}heights}$ pair.

Fish thickness was the maximal distance obtained for each pixel from the difference of the values of the Depth-FR_{aff} and the average value of the background in the $Depth_{aff}$ (Figure 2.6.c).

Five Equidistant Points Thicknesses (Figure 2.A.5.b) were given by the values of $FEP_{3D\text{-}thickness}$.

2.6.4.2 Area (2D) features

All 2D features were given in squared millimeters. In Figure 2.5.c, the main 2D features were shown. Total area was given by the area of the FR_{aff}. The head area, fillet area, ventral area, body-without-caudal-fin area, ventral area, and the fin area (Figure 2.5.c), were given by the area computation of the respective 2D pseudo-landmarks region.

2.6.4.3 Volume (3D) features

All 3D features were given in cubic centimeters. Figure 2.6 shows the three-dimensional representation of the Depth image. For the volume of a fish to be calculated, the integral of the Depth-FR_{aff} must be first determined. This is equivalent to the calculated sum of the gray value g(y,x) of each pixel with coordinates (y,x) belonging the Depth-FR_{aff} , i.e.

$$vol = \sum_{(y,x)\in Depth-FR_{aff}} g(y,x)$$

Following this definition, the total volume of the fish is the grey values area of Depth-FR_{aff}.

To compute the volume for each of the segmented regions the above-described equation was applied to each region (Figure 2.6.a-c). To determine the ventral volume, the integral of Depth-VentralRegion, a cropped region of Depth-FR_{aff} was applied. In the same manner the head volume and fillet volume were computed. The calculation of volume is depicted in Figure 2.7.

Acknowledgement

Authors would like to thank Galaxidi Marine Farm and Cudomar for their participation in the study and their staff for providing necessary help and facilities for raising the population and the experimental data collection.

Funding

This research received funding from the European Commission Horizon 2020 (H2020) Framework Program through grant agreement no. 727315 for the MedAID project (Mediterranean Aquaculture Integrated Development) for data collection of production traits and no. 818367 for the AquaIMPACT project for image acquisition, data analysis and dissemination presented in this manuscript.

Table 2.A.1. Number of records (N), mean and coefficient of variation (CV) of morphometric traits.

	Morphometric trait*	N	Mean	CV
Distance	Total length	931	290.9	5.33
	Standard length	932	250.6	5.53
	Tail excluded length	931	226.4	5.78
	Fillet length	930	162.0	6.45
	Maximum thickness	928	36.3	8.19
	Maximum height	931	103.9	6.84
	1st equidistant height	932	80.8	6.14
	2nd equidistant height	932	102.7	6.67
	3rd equidistant height	932	89.7	6.90
	4th equidistant height	930	59.7	9.71
	5th equidistant height	931	28.7	9.76
	1st equidistant thickness	928	36.8	10.5
	2nd equidistant thickness	921	36.7	8.23
	3rd equidistant thickness	928	31.7	9.38
	4th equidistant thickness	920	25.4	22.8
	Head height	933	90.9	6.21
	Head length	929	64.4	6.35
	Peduncle height	932	22.8	7.18
	Fish eccentricity	928	90.6	0.80
	Head eccentricity	930	62.1	6.10
	Perimeter length	933	1749	6.16
	Perimeter length without tail	926	1547	7.81
	Thickness by height	926	35.0	6.77
	Height by length	930	35.7	4.06
Area	Total area	932	19026	11.2
	Tail area	933	1510	21.9
	Head area	932	3879	11.6
	Relative head area	931	20.4	7.06
	Tail excluded area	933	17523	11.3
	Trunk area	932	13071	12.0
	Fillet area	932	10546	13.9
	Relative fillet area	927	55.4	5.96
	Ventral area	924	2500	23.1
Volume	Total volume	933	494.8	17.7
	Trunk volume	933	355.6	19.0
	Fillet volume	933	281.2	20.4
	Ventral volume	927	73.8	27.4
	Relative ventral volume	927	14.9	21.8
	Ventral volume by fillet volume	927	20.9	22.4
	Head volume	933	139.2	17.1

* All distance and area metrics were in millimeters and squared millimeters, respectively. The volume metrics were cubic centimeters.

Table 2.A.2 Pearson correlation between morphometric measurements.

	Total length	Standard length	Tail excluded length	Fillet length	Maximum thickness	Maximum height	1st equi. height	2nd equi. height	3rd equi. height	4th equi. height	5th equi. height	1st equi. thick.	2nd equi. thick	3rd equi. thick.	4th equi. thick.	Head height	Head length	Peduncle height	Fish eccen.	Head eccen.
Total length	1.000	0.947	0.971	0.933	0.598	0.807	0.880	0.822	0.791	0.649	0.422	0.475	0.628	0.647	0.253	0.841	0.703	0.688	-0.062	0.086
Standard length	0.947	1.000	0.935	0.891	0.562	0.785	0.836	0.797	0.775	0.664	0.464	0.434	0.597	0.618	0.270	0.816	0.690	0.694	0.032	0.048
Tail excluded length	0.971	0.935	1.000	0.959	0.596	0.809	0.866	0.822	0.825	0.738	0.321	0.471	0.622	0.647	0.254	0.847	0.724	0.674	-0.085	0.053
Fillet length	0.933	0.891	0.959	1.000	0.577	0.771	0.856	0.793	0.777	0.680	0.301	0.444	0.610	0.647	0.256	0.757	0.536	0.644	-0.090	0.227
Maximum thickness	0.598	0.562	0.596	0.577	1.000	0.614	0.582	0.604	0.616	0.499	0.369	0.810	0.893	0.645	0.372	0.564	0.414	0.499	-0.220	0.158
Maximum height	0.807	0.785	0.809	0.771	0.614	1.000	0.883	0.986	0.923	0.759	0.446	0.460	0.657	0.738	0.260	0.901	0.597	0.764	-0.504	0.219
1st equidistant height	0.880	0.836	0.866	0.856	0.582	0.883	1.000	0.910	0.864	0.696	0.443	0.450	0.627	0.711	0.274	0.931	0.571	0.775	-0.356	0.366
2nd equidistant height	0.822	0.797	0.822	0.793	0.604	0.986	0.910	1.000	0.930	0.761	0.454	0.456	0.654	0.746	0.266	0.919	0.587	0.788	-0.506	0.250
3rd equidistant height	0.791	0.775	0.825	0.777	0.616	0.923	0.864	0.930	1.000	0.869	0.424	0.461	0.662	0.785	0.299	0.904	0.634	0.802	-0.503	0.178
4th equidistant height	0.649	0.664	0.738	0.680	0.499	0.759	0.696	0.761	0.869	1.000	0.248	0.375	0.522	0.639	0.256	0.761	0.596	0.686	-0.365	0.066
5th equidistant height	0.422	0.464	0.321	0.301	0.369	0.446	0.443	0.454	0.424	0.248	1.000	0.298	0.390	0.461	0.206	0.431	0.249	0.616	-0.032	0.134
1st equidistant thickness	0.475	0.434	0.471	0.444	0.810	0.460	0.450	0.456	0.461	0.375	0.298	1.000	0.718	0.500	0.324	0.442	0.351	0.360	-0.147	0.107
2nd equidistant thickness	0.628	0.597	0.622	0.610	0.893	0.657	0.627	0.654	0.662	0.522	0.390	0.718	1.000	0.706	0.375	0.613	0.437	0.554	-0.262	0.164
3rd equidistant thickness	0.647	0.618	0.647	0.647	0.645	0.738	0.711	0.746	0.785	0.639	0.461	0.500	0.706	1.000	0.345	0.704	0.424	0.705	-0.394	0.225
4th equidistant thickness	0.253	0.270	0.254	0.256	0.372	0.260	0.274	0.266	0.299	0.256	0.206	0.324	0.375	0.345	1.000	0.259	0.157	0.314	-0.072	0.094
Head height	0.841	0.816	0.847	0.757	0.564	0.901	0.931	0.919	0.904	0.761	0.431	0.442	0.613	0.704	0.259	1.000	0.752	0.786	-0.415	0.150
Head length	0.703	0.690	0.724	0.536	0.414	0.597	0.571	0.587	0.634	0.596	0.249	0.351	0.437	0.424	0.157	0.752	1.000	0.494	-0.049	-0.408
Peduncle height	0.688	0.694	0.674	0.644	0.499	0.764	0.775	0.788	0.802	0.686	0.616	0.360	0.554	0.705	0.314	0.786	0.494	1.000	-0.307	0.171
Fish eccentricity	-0.062	0.032	-0.085	-0.090	-0.220	-0.504	-0.356	-0.506	-0.503	-0.365	-0.032	-0.147	-0.262	-0.394	-0.072	-0.415	-0.049	-0.307	1.000	-0.364
Head eccentricity	0.086	0.048	0.053	0.227	0.158	0.219	0.366	0.250	0.178	0.066	0.134	0.107	0.164	0.225	0.094	0.150	-0.408	0.171	-0.364	1.000
Perimeter length	0.857	0.860	0.836	0.800	0.597	0.740	0.743	0.736	0.723	0.616	0.433	0.469	0.610	0.590	0.248	0.715	0.614	0.604	-0.022	0.078
Perimeter length without tail	0.659	0.681	0.650	0.611	0.486	0.591	0.598	0.583	0.566	0.503	0.390	0.398	0.474	0.465	0.187	0.591	0.511	0.510	-0.038	0.062
Thickness by height	-0.095	-0.119	-0.096	-0.079	0.599	-0.261	-0.189	-0.259	-0.188	-0.163	-0.012	0.490	0.424	0.022	0.191	-0.233	-0.108	-0.172	0.240	-0.038
Height by length	0.043	0.078	0.084	0.072	0.246	0.624	0.329	0.572	0.514	0.420	0.195	0.150	0.279	0.387	0.103	0.412	0.076	0.379	-0.790	0.254
Total area	0.947	0.928	0.945	0.908	0.642	0.915	0.937	0.931	0.924	0.797	0.496	0.496	0.682	0.749	0.298	0.923	0.685	0.806	-0.256	0.175
Tail area	0.488	0.513	0.469	0.455	0.283	0.496	0.495	0.515	0.506	0.399	0.401	0.182	0.323	0.427	0.179	0.490	0.319	0.469	-0.132	0.097
Head area	0.800	0.785	0.814	0.653	0.504	0.752	0.774	0.751	0.775	0.688	0.336	0.409	0.540	0.559	0.224	0.900	0.944	0.646	-0.189	-0.156
Relative head area	-0.215	-0.214	-0.190	-0.405	-0.219	-0.244	-0.240	-0.276	-0.230	-0.163	-0.255	-0.140	-0.221	-0.301	-0.112	-0.015	0.454	-0.248	0.113	-0.576
Tail excluded area	0.944	0.920	0.945	0.906	0.651	0.910	0.932	0.922	0.917	0.796	0.476	0.509	0.685	0.742	0.292	0.919	0.691	0.794	-0.259	0.174
Trunk area	0.930	0.894	0.935	0.941	0.637	0.908	0.929	0.925	0.914	0.793	0.441	0.488	0.674	0.752	0.289	0.877	0.582	0.780	-0.293	0.257
Fillet area	0.850	0.820	0.848	0.851	0.463	0.835	0.889	0.867	0.846	0.716	0.417	0.333	0.532	0.685	0.235	0.847	0.535	0.737	-0.303	0.257
Relative fillet area	0.215	0.172	0.212	0.284	-0.120	0.242	0.325	0.282	0.246	0.184	0.032	-0.150	-0.044	0.202	0.005	0.256	-0.025	0.211	-0.232	0.281
Ventral area	0.344	0.332	0.362	0.367	0.552	0.307	0.242	0.283	0.313	0.313	0.136	0.481	0.493	0.286	0.168	0.215	0.207	0.231	0.001	0.046
Total volume	0.889	0.844	0.887	0.856	0.809	0.867	0.872	0.872	0.884	0.747	0.477	0.649	0.815	0.818	0.353	0.854	0.640	0.745	-0.289	0.188
Trunk volume	0.879	0.828	0.876	0.878	0.774	0.870	0.877	0.880	0.886	0.743	0.477	0.609	0.796	0.847	0.347	0.831	0.558	0.754	-0.319	0.257
Fillet volume	0.862	0.813	0.856	0.854	0.651	0.854	0.885	0.875	0.874	0.725	0.469	0.505	0.705	0.825	0.308	0.846	0.559	0.751	-0.336	0.252
Ventral volume	0.464	0.443	0.472	0.483	0.711	0.449	0.392	0.430	0.454	0.400	0.251	0.585	0.652	0.465	0.263	0.352	0.267	0.363	-0.088	0.123
Relative ventral volume	-0.133	-0.118	-0.116	-0.074	0.240	-0.126	-0.204	-0.150	-0.130	-0.086	-0.067	0.198	0.174	-0.073	0.049	-0.239	-0.181	-0.131	0.105	0.010
Ventral volume by fillet volume	-0.178	-0.156	-0.160	-0.145	0.216	-0.182	-0.259	-0.211	-0.186	-0.129	-0.095	0.187	0.135	-0.150	0.030	-0.266	-0.145	-0.184	0.148	-0.062
Head volume	0.790	0.765	0.789	0.670	0.793	0.732	0.731	0.721	0.750	0.646	0.408	0.668	0.751	0.616	0.318	0.798	0.786	0.613	-0.163	-0.035

Perim. length	Perim. length without tail	Thickness by height	Height by length	Total area	Tail area	Head area	Relative head area	Tail excluded area	Trunk area	Fillet area	Relative fillet area	Ventral area	Total volume	Trunk volume	Fillet volume	Ventral volume	Relative ventral volume	Ventral volume by fillet volume	Head volume
0.857	0.659	-0.095	0.043	0.947	0.488	0.800	-0.215	0.944	0.930	0.850	0.215	0.344	0.889	0.879	0.862	0.464	-0.133	-0.178	0.790
0.860	0.681	-0.119	0.078	0.928	0.513	0.785	-0.214	0.920	0.894	0.820	0.172	0.332	0.844	0.828	0.813	0.443	-0.118	-0.156	0.765
0.836	0.650	-0.096	0.084	0.945	0.469	0.814	-0.190	0.945	0.935	0.848	0.212	0.362	0.887	0.876	0.856	0.472	-0.116	-0.160	0.789
0.800	0.611	-0.079	0.072	0.908	0.455	0.653	-0.405	0.906	0.941	0.851	0.284	0.367	0.856	0.878	0.854	0.483	-0.074	-0.145	0.670
0.597	0.486	0.599	0.246	0.642	0.283	0.504	-0.219	0.651	0.637	0.463	-0.120	0.552	0.809	0.774	0.651	0.711	0.240	0.216	0.793
0.740	0.591	-0.261	0.624	0.915	0.496	0.752	-0.244	0.910	0.908	0.835	0.242	0.307	0.867	0.870	0.854	0.449	-0.126	-0.182	0.732
0.743	0.598	-0.189	0.329	0.937	0.495	0.774	-0.240	0.932	0.929	0.889	0.325	0.242	0.872	0.877	0.885	0.392	-0.204	-0.259	0.731
0.736	0.583	-0.259	0.572	0.931	0.515	0.751	-0.276	0.922	0.925	0.867	0.282	0.283	0.872	0.880	0.875	0.430	-0.150	-0.211	0.721
0.723	0.566	-0.188	0.514	0.924	0.506	0.775	-0.230	0.917	0.914	0.846	0.246	0.313	0.884	0.886	0.874	0.454	-0.130	-0.186	0.750
0.616	0.503	-0.163	0.420	0.797	0.399	0.688	-0.163	0.796	0.793	0.716	0.184	0.313	0.747	0.743	0.725	0.400	-0.086	-0.129	0.646
0.433	0.390	-0.012	0.195	0.496	0.401	0.336	-0.255	0.476	0.441	0.417	0.032	0.136	0.477	0.477	0.469	0.251	-0.067	-0.095	0.408
0.469	0.398	0.490	0.150	0.496	0.182	0.409	-0.140	0.509	0.488	0.333	-0.150	0.481	0.649	0.609	0.505	0.585	0.198	0.187	0.668
0.610	0.474	0.424	0.279	0.682	0.323	0.540	-0.221	0.685	0.674	0.532	-0.044	0.493	0.815	0.796	0.705	0.652	0.174	0.135	0.751
0.590	0.465	0.022	0.387	0.749	0.427	0.559	-0.301	0.742	0.752	0.685	0.202	0.286	0.818	0.847	0.825	0.465	-0.073	-0.150	0.616
0.248	0.187	0.191	0.103	0.298	0.179	0.224	-0.112	0.292	0.289	0.235	0.005	0.168	0.353	0.347	0.308	0.263	0.049	0.030	0.318
0.715	0.591	-0.233	0.412	0.923	0.490	0.900	-0.015	0.919	0.877	0.847	0.256	0.215	0.854	0.831	0.846	0.352	-0.239	-0.266	0.798
0.614	0.511	-0.108	0.076	0.685	0.319	0.944	0.454	0.691	0.582	0.535	-0.025	0.207	0.640	0.558	0.559	0.267	-0.181	-0.145	0.786
0.604	0.510	-0.172	0.379	0.806	0.469	0.646	-0.248	0.794	0.780	0.737	0.211	0.231	0.745	0.754	0.751	0.363	-0.131	-0.184	0.613
-0.022	-0.038	0.240	-0.790	-0.256	-0.132	-0.189	0.113	-0.259	-0.293	-0.303	-0.232	0.001	-0.289	-0.319	-0.336	-0.088	0.105	0.148	-0.163
0.078	0.062	-0.038	0.254	0.175	0.097	-0.156	-0.576	0.174	0.257	0.257	0.281	0.046	0.188	0.257	0.252	0.123	0.010	-0.062	-0.035
1.000	0.619	-0.033	0.118	0.860	0.509	0.690	-0.265	0.848	0.825	0.718	0.071	0.409	0.818	0.806	0.764	0.507	-0.021	-0.063	0.736
0.619	1.000	-0.006	0.119	0.660	-0.011	0.584	-0.114	0.721	0.643	0.526	-0.007	0.379	0.630	0.612	0.551	0.450	0.058	0.032	0.592
-0.033	-0.006	1.000	-0.318	-0.149	-0.166	-0.154	-0.014	-0.136	-0.145	-0.289	-0.401	0.362	0.094	0.050	-0.084	0.404	0.429	0.457	0.208
0.118	0.119	-0.318	1.000	0.292	0.187	0.211	-0.124	0.289	0.303	0.283	0.123	0.060	0.288	0.307	0.303	0.142	-0.040	-0.073	0.192
0.860	0.660	-0.149	0.292	1.000	0.553	0.823	-0.268	0.990	0.980	0.901	0.233	0.350	0.939	0.936	0.920	0.488	-0.133	-0.188	0.812
0.509	-0.011	-0.166	0.187	0.553	1.000	0.401	-0.247	0.438	0.491	0.496	0.109	0.091	0.489	0.489	0.516	0.179	-0.169	-0.197	0.416
0.690	0.584	-0.154	0.211	0.823	0.401	1.000	0.320	0.826	0.731	0.692	0.075	0.218	0.765	0.698	0.705	0.317	-0.217	-0.200	0.849
-0.265	-0.114	-0.014	-0.124	-0.268	-0.247	0.320	1.000	-0.249	-0.386	-0.335	-0.268	-0.214	-0.268	-0.376	-0.339	-0.277	-0.147	-0.024	0.078
0.848	0.721	-0.136	0.289	0.990	0.438	0.826	-0.249	1.000	0.980	0.894	0.235	0.366	0.937	0.933	0.911	0.502	-0.118	-0.172	0.813
0.825	0.643	-0.145	0.303	0.980	0.491	0.731	-0.386	0.980	1.000	0.917	0.302	0.359	0.930	0.948	0.930	0.500	-0.113	-0.187	0.741
0.718	0.526	-0.289	0.283	0.901	0.496	0.692	-0.335	0.894	0.917	1.000	0.615	-0.003	0.807	0.837	0.939	0.163	-0.451	-0.520	0.606
0.071	-0.007	-0.401	0.123	0.233	0.109	0.075	-0.268	0.235	0.302	0.615	1.000	-0.683	0.128	0.198	0.442	-0.560	-0.843	-0.896	-0.089
0.409	0.379	0.362	0.060	0.350	0.091	0.218	-0.214	0.366	0.359	-0.003	-0.683	1.000	0.452	0.428	0.171	0.955	0.861	0.828	0.453
0.818	0.630	0.094	0.288	0.939	0.489	0.765	-0.268	0.937	0.930	0.807	0.128	0.452	1.000	0.987	0.930	0.622	-0.020	-0.071	0.892
0.806	0.612	0.050	0.307	0.936	0.489	0.698	-0.376	0.933	0.948	0.837	0.198	0.428	0.987	1.000	0.953	0.602	-0.033	-0.106	0.808
0.764	0.551	-0.084	0.303	0.920	0.516	0.705	-0.339	0.911	0.930	0.939	0.442	0.171	0.930	0.953	1.000	0.352	-0.305	-0.376	0.732
0.507	0.450	0.404	0.142	0.488	0.179	0.317	-0.277	0.502	0.500	0.163	-0.560	0.955	0.622	0.602	0.352	1.000	0.756	0.714	0.590
-0.021	0.058	0.429	-0.040	-0.133	-0.169	-0.217	-0.147	-0.118	-0.113	-0.451	-0.843	0.861	-0.020	-0.033	-0.305	0.756	1.000	0.987	0.019
-0.063	0.032	0.457	-0.073	-0.188	-0.197	-0.200	-0.024	-0.172	-0.187	-0.520	-0.896	0.828	-0.071	-0.106	-0.376	0.714	0.987	1.000	0.039
0.736	0.592	0.208	0.192	0.812	0.416	0.849	0.078	0.813	0.741	0.606	-0.089	0.453	0.892	0.808	0.732	0.590	0.019	0.039	1.000

Table 2.A.3. Genetic and environmental variance, and heritability estimates of morphometric traits.

	Morphometric trait*	V_A	V_E	h^2 ± se
Distance	Total length	16.95	121.77	0.44 ± 0.05
	Standard length	14.18	29.90	0.32 ± 0.05
	Tail excluded length	13.22	14.35	0.48 ± 0.05
	Fillet length	10.89	13.97	0.44 ± 0.05
	Maximum thickness	0.85	2.80	0.23 ± 0.05
	Maximum height	1.59	7.45	0.17 ± 0.05
	1st equidistant height	0.98	2.16	0.31 ± 0.05
	2nd equidistant height	2.19	4.94	0.31 ± 0.05
	3rd equidistant height	1.97	3.96	0.33 ± 0.05
	4th equidistant height	3.58	9.61	0.27 ± 0.05
	5th equidistant height	2.03	4.08	0.33 ± 0.05
	1st equidistant thickness	0.97	8.25	0.11 ± 0.05
	2nd equidistant thickness	0.60	3.06	0.16 ± 0.05
	3rd equidistant thickness	0.43	2.04	0.17 ± 0.05
	4th equidistant thickness	1.26	28.98	0.04 ± 0.03
	Head height	1.81	3.44	0.35 ± 0.06
	Head length	2.51	5.96	0.30 ± 0.05
	Peduncle height	0.30	0.56	0.35 ± 0.05
	Fish eccentricity	$1.5 * 10^{-5}$	$2.9 * 10^{-5}$	0.35 ± 0.05
	Head eccentricity	$3.4 * 10^{-4}$	$9.0 * 10^{-4}$	0.27 ± 0.05
	Perimeter length	842	2515	0.25 ± 0.05
	Perimeter length without tail	945	7697	0.11 ± 0.05
	Thickness by height	$1.2 * 10^{-4}$	$3.1 * 10^{-4}$	0.28 ± 0.05
	Height by length	$5.3 * 10^{-5}$	$1.2 * 10^{-4}$	0.30 ± 0.05
Area	Total area	90768	152323	0.37 ± 0.05
	Tail area	5088	72336	0.06 ± 0.05
	Head area	22391	49169	0.31 ± 0.05
	Relative head area	0.40	1.39	0.22 ± 0.05
	Tail excluded area	77139	185844	0.29 ± 0.05
	Trunk area	43119	131874	0.25 ± 0.05
	Fillet area	60858	347050	0.15 ± 0.05
	Relative fillet area	1.00	7.80	0.11 ± 0.04
	Ventral area	52228	181734	0.22 ± 0.05
Volume	Total volume	28.91	406.0	0.07 ± 0.03
	Trunk volume	28.45	204.94	0.12 ± 0.04
	Fillet volume	24.69	302.20	0.07 ± 0.04
	Ventral volume	56.31	186.14	0.23 ± 0.05
	Relative ventral volume	2.37	6.99	0.25 ± 0.05
	Ventral volume by fillet volume	4.39	14.41	0.23 ± 0.05
	Head volume	28.98	138.84	0.23 ± 0.05

* All distance and area metrics were in millimeters and squared millimeters, respectively. The volume metrics were cubic centimeters.

Table 2.A.4. Genetic correlations between morphometric traits and production traits.

	Morphometric trait*	HW#	FW	FF	VisW
Distance	Total length	0.93 ± 0.02	-0.30 ± 0.19	-0.19 ± 0.10	-0.24 ± 0.11
	Standard length	0.92 ± 0.02	-0.39 ± 0.19	-0.34 ± 0.11	-0.32 ± 0.11
	Tail excluded length	0.92 ± 0.02	-0.29 ± 0.18	-0.26 ± 0.10	-0.29 ± 0.10
	Fillet length	0.90 ± 0.02	0.16 ± 0.19	0.02 ± 0.11	-0.17 ± 0.11
	Maximum thickness	0.84 ± 0.04	0.28 ± 0.22	0.42 ± 0.12	-0.02 ± 0.14
	Maximum height	0.97 ± 0.01	-0.07 ± 0.25	-0.25 ± 0.15	-0.18 ± 0.16
	1st equidistant height	0.96 ± 0.01	-0.25 ± 0.20	-0.30 ± 0.11	-0.20 ± 0.12
	2nd equidistant height	0.96 ± 0.01	-0.13 ± 0.21	-0.23 ± 0.12	-0.12 ± 0.14
	3rd equidistant height	0.96 ± 0.01	-0.21 ± 0.20	-0.20 ± 0.12	-0.07 ± 0.12
	4th equidistant height	0.90 ± 0.03	-0.13 ± 0.21	-0.27 ± 0.12	-0.13 ± 0.13
	5th equidistant height	0.59 ± 0.08	0.07 ± 0.21	0.06 ± 0.12	0.13 ± 0.12
	1st equidistant thickness	0.82 ± 0.07	0.21 ± 0.28	0.55 ± 0.18	-0.07 ± 0.18
	2nd equidistant thickness	0.91 ± 0.03	0.42 ± 0.22	0.63 ± 0.13	0.13 ± 0.15
	3rd equidistant thickness	0.96 ± 0.01	0.61 ± 0.17	0.72 ± 0.09	0.19 ± 0.15
	4th equidistant thickness	0.80 ± 0.12	-0.15 ± 0.38	0.60 ± 0.37	0.31 ± 0.34
	Head height	0.95 ± 0.01	-0.63 ± 0.14	-0.61 ± 0.09	-0.27 ± 0.12
	Head length	0.78 ± 0.05	-0.68 ± 0.13	-0.57 ± 0.10	-0.25 ± 0.12
	Peduncle height	0.90 ± 0.02	0.07 ± 0.21	-0.12 ± 0.11	-0.27 ± 0.12
	Fish eccentricity	-0.56 ± 0.08	-0.18 ± 0.20	-0.07 ± 0.11	-0.13 ± 0.12
	Head eccentricity	0.31 ± 0.12	0.57 ± 0.19	0.16 ± 0.12	0.06 ± 0.13
	Perimeter length	0.90 ± 0.03	0.02 ± 0.23	-0.22 ± 0.13	-0.34 ± 0.13
	Perimeter length without tail	0.86 ± 0.05	-0.15 ± 0.27	-0.15 ± 0.16	-0.27 ± 0.17
	Thickness by height	-0.31 ± 0.19	0.23 ± 0.20	0.42 ± 0.12	-0.01 ± 0.13
	Height by length	0.55 ± 0.09	0.14 ± 0.21	0.02 ± 0.12	0.10 ± 0.13
Area	Total area	0.98 ± 0.01	-0.31 ± 0.19	-0.52 ± 0.10	-0.36 ± 0.11
	Tail area	0.89 ± 0.05	-0.29 ± 0.31	-0.34 ± 0.21	-0.15 ± 0.22
	Head area	0.88 ± 0.03	-0.69 ± 0.13	-0.61 ± 0.09	-0.25 ± 0.12
	Relative head area	-0.37 ± 0.12	-0.51 ± 0.19	-0.58 ± 0.10	-0.13 ± 0.14
	Tail excluded area	0.98 ± 0.01	-0.25 ± 0.20	-0.47 ± 0.11	-0.37 ± 0.12
	Trunk area	0.98 ± 0.01	0.27 ± 0.22	-0.23 ± 0.13	-0.26 ± 0.14
	Fillet area	0.97 ± 0.01	-0.44 ± 0.24	-0.75 ± 0.15	0.04 ± 0.16
	Relative fillet area	0.55 ± 0.13	-0.77 ± 0.12	-0.54 ± 0.11	0.33 ± 0.16
	Ventral area	0.61 ± 0.10	0.60 ± 0.22	0.39 ± 0.13	-0.20 ± 0.14
Volume	Total volume	1.00 ± 0.002	0.45 ± 0.28	-0.05 ± 0.20	-0.37 ± 0.20
	Trunk volume	0.99 ± 0.003	0.89 ± 0.10	0.45 ± 0.16	-0.04 ± 0.17
	Fillet volume	0.99 ± 0.003	-0.10 ± 0.35	-0.31 ± 0.22	0.23 ± 0.20
	Ventral volume	0.70 ± 0.07	0.64 ± 0.20	0.43 ± 0.13	-0.18 ± 0.13
	Relative ventral volume	-0.24 ± 0.13	0.62 ± 0.21	0.45 ± 0.13	-0.16 ± 0.13
	Ventral volume by fillet volume	-0.35 ± 0.12	0.52 ± 0.23	0.39 ± 0.14	-0.20 ± 0.13
	Head volume	0.93 ± 0.02	-0.54 ± 0.20	-0.47 ± 0.12	-0.36 ± 0.13

HW = Harvest weight, FW = Fillet weight, FF = Fillet fat percentage, VisW = Viscera weight

Harvest weight was a covariate in the genetic models for all traits except for harvest weight

* All distance and area metrics were in millimeters and squared millimeters, respectively. The volume metrics were cubic centimeters.

Table 2.A.5. Phenotypic correlations between morphometric traits and production traits.

	Morphometric trait*	HW	FW	FF	VisW
Distance	Total length	0.92 ± 0.01	0.87 ± 0.01	0.20 ± 0.04	0.70 ± 0.02
	Standard length	0.88 ± 0.01	0.83 ± 0.01	0.18 ± 0.04	0.67 ± 0.02
	Tail excluded length	0.92 ± 0.01	0.87 ± 0.01	0.20 ± 0.04	0.69 ± 0.02
	Fillet length	0.89 ± 0.01	0.86 ± 0.01	0.24 ± 0.04	0.67 ± 0.02
	Maximum thickness	0.73 ± 0.02	0.72 ± 0.02	0.27 ± 0.04	0.62 ± 0.03
	Maximum height	0.91 ± 0.01	0.87 ± 0.01	0.25 ± 0.04	0.74 ± 0.02
	1st equidistant height	0.94 ± 0.01	0.89 ± 0.01	0.19 ± 0.04	0.72 ± 0.02
	2nd equidistant height	0.93 ± 0.01	0.88 ± 0.01	0.27 ± 0.04	0.75 ± 0.02
	3rd equidistant height	0.93 ± 0.01	0.88 ± 0.01	0.26 ± 0.04	0.75 ± 0.02
	4th equidistant height	0.79 ± 0.02	0.75 ± 0.02	0.17 ± 0.04	0.62 ± 0.03
	5th equidistant height	0.51 ± 0.03	0.50 ± 0.03	0.13 ± 0.04	0.44 ± 0.03
	1st equidistant thickness	0.55 ± 0.03	0.55 ± 0.03	0.18 ± 0.04	0.44 ± 0.03
	2nd equidistant thickness	0.77 ± 0.02	0.75 ± 0.02	0.30 ± 0.04	0.65 ± 0.02
	3rd equidistant thickness	0.87 ± 0.01	0.86 ± 0.01	0.40 ± 0.04	0.75 ± 0.02
	4th equidistant thickness	0.34 ± 0.03	0.35 ± 0.03	0.16 ± 0.04	0.28 ± 0.03
	Head height	0.92 ± 0.01	0.86 ± 0.01	0.15 ± 0.04	0.71 ± 0.02
	Head length	0.68 ± 0.02	0.60 ± 0.03	-0.03 ± 0.04	0.49 ± 0.03
	Peduncle height	0.84 ± 0.01	0.81 ± 0.01	0.20 ± 0.04	0.66 ± 0.03
	Fish eccentricity	-0.38 ± 0.04	-0.36 ± 0.04	-0.23 ± 0.04	-0.37 ± 0.04
	Head eccentricity	0.25 ± 0.04	0.25 ± 0.04	0.13 ± 0.04	0.19 ± 0.04
	Perimeter length	0.83 ± 0.01	0.79 ± 0.02	0.24 ± 0.04	0.65 ± 0.03
	Perimeter length without tail	0.63 ± 0.02	0.60 ± 0.03	0.11 ± 0.04	0.47 ± 0.03
	Thickness by height	-0.14 ± 0.04	-0.10 ± 0.04	0.14 ± 0.04	-0.09 ± 0.04
	Height by length	0.36 ± 0.04	0.35 ± 0.04	0.23 ± 0.04	0.38 ± 0.04
Area	Total area	0.98 ± 0.002	0.92 ± 0.01	0.24 ± 0.04	0.77 ± 0.02
	Tail area	0.55 ± 0.03	0.53 ± 0.03	0.18 ± 0.04	0.46 ± 0.03
	Head area	0.81 ± 0.02	0.73 ± 0.02	0.06 ± 0.04	0.60 ± 0.03
	Relative head area	-0.30 ± 0.04	-0.35 ± 0.04	-0.37 ± 0.03	-0.27 ± 0.04
	Tail excluded area	0.97 ± 0.003	0.91 ± 0.01	0.23 ± 0.04	0.76 ± 0.02
	Trunk area	0.97 ± 0.003	0.92 ± 0.01	0.27 ± 0.04	0.77 ± 0.02
	Fillet area	0.91 ± 0.01	0.86 ± 0.01	0.23 ± 0.04	0.74 ± 0.02
	Relative fillet area	0.32 ± 0.04	0.30 ± 0.04	0.003 ± 0.04	0.29 ± 0.04
	Ventral area	0.37 ± 0.04	0.39 ± 0.04	0.17 ± 0.04	0.27 ± 0.04
Volume	Total volume	0.97 ± 0.002	0.94 ± 0.01	0.30 ± 0.04	0.79 ± 0.02
	Trunk volume	0.97 ± 0.002	0.95 ± 0.004	0.33 ± 0.04	0.80 ± 0.02
	Fillet volume	0.95 ± 0.004	0.92 ± 0.01	0.30 ± 0.04	0.79 ± 0.02
	Ventral volume	0.53 ± 0.03	0.50 ± 0.03	0.23 ± 0.04	0.38 ± 0.04
	Relative ventral volume	-0.18 ± 0.04	-0.15 ± 0.04	0.03 ± 0.04	-0.17 ± 0.04
	Ventral volume by fillet volume	-0.25 ± 0.04	-0.23 ± 0.04	-0.03 ± 0.04	-0.23 ± 0.04
	Head volume	0.83 ± 0.01	0.74 ± 0.02	0.14 ± 0.04	0.64 ± 0.03

HW = Harvest weight, FW = Fillet weight, FF = Fillet fat percentage, VisW = Viscera weight

* All distance and area metrics were in millimeters and squared millimeters, respectively. The volume metrics were cubic centimeters.

Figure 2.A.1. X-Rite ColorChecker® Classic with 24 natural objects.

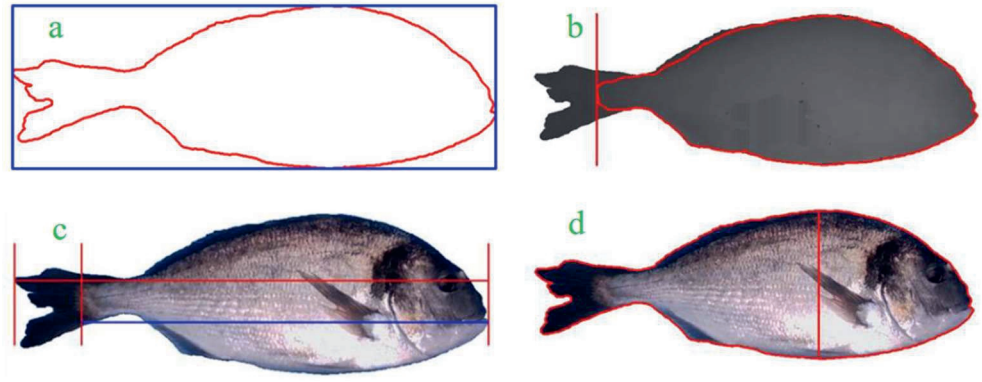

Figure 2.A.2. a) the smallest rectangle orthogonal to the image reference system (IRS). b) the contour of rough segmentation of the body without caudal fin and the vertical line passing through the tip of the body. c) the total length (red line) and the standard length (blue line) of the fish. d) the maximum height of the fish.

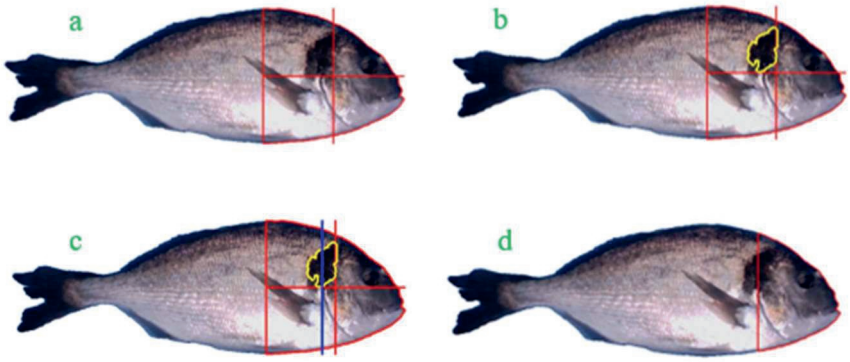

Figure 2.A.3. a) the front region and the four quadrants in which this region was automatically divided. b) a large part of the black spot was in the first quadrant. c) the black spot centroid defines the vertical line that separate the head region from the rest of the body (blue line). d) the head region.

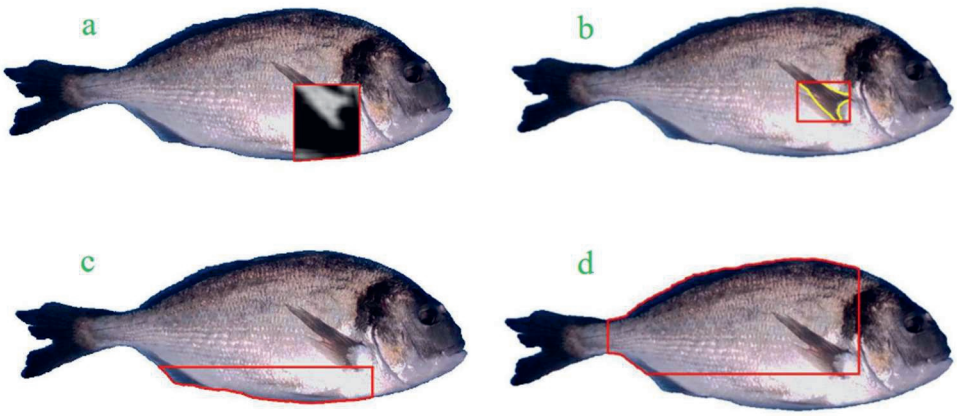

Figure 2.A.4. a) the entropy of the gray values of the blue channel image limited to fourth quadrant. b) the base of the lateral fin region and the smallest rectangle around this region. c) the ventral region. d) the fillet region.

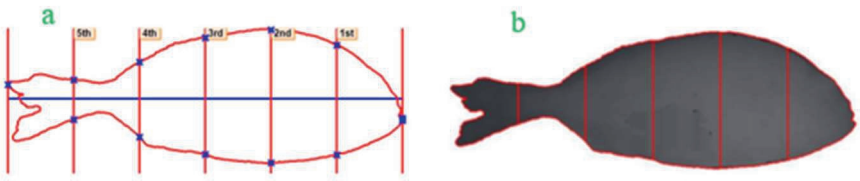

Figure 2.A.5. The automatic computation of the five equidistant points (FEP) and their projection on the segmented fish in a) RGB image and in b) Depth image. The $FEP_{3D\text{-}thickness}$ are thickness values, expressed in millimeters, of the fish evaluated using the depth image onto each equidistant point.

Chapter 3:

A Novel Approach for Genomic Selection in Gilthead Seabream (*Sparus aurata*) Using Image Analysis to Predict the Phenotypes of Selection Candidates

Benan Gulzari[1], Angelo Mencarelli[2], Mark Camara[1], Chantal Roozeboom[1], Hans Komen[1], John W.M. Bastiaansen[1]

[1] Department of Animal Breeding and Genomics, Wageningen University and Research, Droevendaalsesteeg 1, 6700 AH, Wageningen (The Netherlands)

[2] Wageningen Plant Research, Wageningen University and Research, Droevendaalsesteeg 1, 6700 AH, Wageningen (The Netherlands)

Abstract

Genomic selection (GS) accelerates genetic gain compared to classical selection methods but is not widespread in fish breeding in part due to the high cost of genotyping a large reference population. This cost can be eliminated entirely using computer vision to predict phenotypes for selection candidates and estimating breeding values using predicted phenotypes for the selection candidates without genotyping a reference population. Instead, images and observed phenotypes provide a training dataset for phenotypic prediction models from the reference population. This paper compares the accuracy of GEBVs between GS designs where the reference population contributes genotypes and phenotypes (i.e., standard genomic selection) (GS_S), and a reference population that contributes phenotypic prediction equations without genotypes and genomic analyses use predicted phenotypes for selection candidates' own predicted records (genomic selection with predicted phenotypes) (GS_PP). For a number of traits, the GEBVs were more accurate in GS_PP designs than in GS_S designs with less genotyping effort and cost (8% for harvest weight, 13% for fillet weight, 44% for fillet yield, 1% for fillet fat, 10% for viscera weight, and 1% for viscerosomatic index). Our findings show that image analysis can reduce the genotyping costs of a GS breeding program and simultaneously improve the accuracy of the GEBVs.

3.1 Introduction

Genomic selection (GS) accelerates genetic gain compared to classical selection methods (Meuwissen et al., 2001), and has become routine for terrestrial farm animals in the last two decades. However, GS is applied routinely only in Atlantic salmon (*Salmo salar*) (Boudry et al., 2021). For other farmed fish species, GS is used experimentally, or its application is still in early stages (Zenger et al., 2017).

In standard GS breeding programs, selection candidates and performance tested animals in the reference population must be genotyped to estimate breeding values for the candidates via the genomic relationship matrix (GRM) (Kriaridou et al., 2020). A large reference population may be necessary to obtain high accuracies, especially for traits with low heritabilities (Goddard et al., 2011), incurring high costs for implementation and impeding the adoption of this technology. Much research has focused on reducing genotyping costs including lowering the density of SNP panels (Kriaridou et al., 2020), imputation of SNPs (Bolormaa et al., 2015), and reducing sib testing (Sonesson and Meuwissen, 2009).

In fish breeding programs, breeding goals may consist of sib traits that require slaughtering fish to be quantified, such as processing yield and product quality (Chavanne et al., 2016). If included in the breeding goals, these traits are improved using sib data. Selection on sib data limits genetic progress because only the between family variance is exploited unless genomic selection is used to capture the within family Mendelian sampling variance. Alternatively, indicator traits measured on selection candidates can be used to genetically improve sib traits. However, the efficiency of selection using indicator traits depends on both the genetic correlation with the sib traits and heritability (Falconer and Mackay, 1996). In Chapter 2, we showed that computer vision yields several indicator traits that have high heritabilities and genetic correlations with important sib traits.

Computer vision involves extraction of information from digital images (Dowlati et al., 2012; Navarro et al., 2016). In Chapter 2, we developed an automated

process to extract morphometric measurements from 2-dimensional (2D) and 3-dimensional (3D) images of gilthead seabream. The automated computer vision process yielded several morphometric measurements that consisted of distances, areas, and volumes of different body parts. In this study, we aim to establish a novel low-cost GS design using the morphometric measurements produced in Chapter 2. This GS design uses predicted phenotypes of selection candidates as performance data. In this design, ungenotyped fish contributes phenotypes for sib traits and automated morphometric measurements. These phenotypes and morphometric measurements are then used to build phenotypic prediction models and predict the phenotypes of selection candidates for sib traits. The proposed GS design eliminates the need to genotype a reference population, which results in reduced costs compared to traditional GS designs.

To achieve this, the objectives of this paper are to i) predict phenotypes of production traits using morphometric measurements in gilthead seabream that were obtained in Chapter 2, ii) estimate the genetic correlations between observed and predicted phenotypes, and finally iii) compare the accuracy of GEBVs between standard GS designs (GS_S) and a novel design where genetic analyses are performed using predicted phenotypes for selection candidates (genomic selection with predicted phenotypes) (GS_PP).

3.2 Materials & Methods

3.2.1 Production sites and production of experimental fish

For this experiment, a population of juveniles was produced by mass spawning 33 males and 20 females on a single day. When fish reached an average weight of 3 grams, a batch of 84,650 juveniles was stocked in a sea cage near El Campello, Spain (Cudomar SL) and grown under commercial conditions.

3.2.2 Phenotypic data collection

We collected data on production traits from the commercially produced fish at harvest after a grow-out period of 500 - 504 days. In total, we measured 945 fish over 5 days (daily 150 – 200 fish). The data collection process was explained in

detail in (Gulzari et al., 2022). In short, we collected data on 1) harvest weight, 2) fillet fat percentage estimations on whole fish from eight points (four on each side of the fish) using a Distell Fish Fat meter (Distell Inc., West Lothian, Scotland) and averaging the eight measurements on the same fish, 3) viscera weight (including all the internal organs and fat), 4) viscera percentage as (viscera weight / harvest weight) x 100, 5) fillet weight for only left side and multiplied by two to estimate the total fillet weight, 6) fillet yield as (total fillet weight/harvest weight) x 100. Measurements taken on a single day were performed by the same person for each trait.

We used R software and the "tidyverse" package collection to edit and organize the data (R Core Team, 2020; Wickham et al., 2019).

3.2.3 Image acquisition and morphometric measurements

The 2D and 3D images of the whole fish were acquired using an Intel® RealSense™ D435 Depth Camera (Intel Corp., Santa Clara, California, USA), we automated the extraction of morphometric measurements with Halcon 20.11 framework (MVTec Software GmbH, Munich, Germany) (Eckstein and Steger, 1999). The automated process of morphometric measurement extraction was explained in detail in Chapter 2. In short, 40 morphometric measurements were extracted from 2D and 3D images of 933 gilthead seabream including 24 distance, nine area, and seven volume measurements.

3.2.4 Prediction of phenotypes and genetic analyses

3.2.4.1 Prediction of phenotypes

We used multiple linear regression models to create phenotypic prediction models using all available data. In these models, we defined the phenotypic prediction accuracy as the correlation between observed and predicted phenotypes, which was tested using 10-fold cross-validation. The predictor variables in those models were the morphometric measurements and their interactions. We added predictors to the linear regression models using forward

selection. The models were called “final models” when they resulted in the highest correlation between observed and predicted phenotypes.

3.2.4.2 Genetic analyses

3.2.4.2.1 DNA extraction and genomic relationship matrix

IdentiGEN (Dublin, Ireland) isolated the DNA from fin clips and performed the genotyping using the ~30k “MedFish” SNP array (Peñaloza et al., 2021). We excluded SNPs that had no call rates exceeding 10%, that were monomorphic, or that had Hardy-Weinberg equilibrium exact test p-values below 1e-10. We computed a genomic relationship matrix (GRM) based on the remaining 28164 SNPs by using “vanraden” option (VanRaden, 2008) in calc_grm software (Calus and Vandenplas, 2019), in which the GRM is computed as $G = \frac{ZZ'}{2\sum p_i p_i (1-p_i)}$. To use this option, we coded marker genotypes as “0”, “1”, or “2”. Z is a matrix that contains marker genotypes for all loci, corrected for the allele frequencies per locus. p_i is the frequency of the minor allele. The inverse of the GRM matrix was obtained directly from calc_grm by using “giv” function.

3.2.4.2.2 Estimation of heritabilities and correlations

We estimated the heritabilities of and genetic correlations between observed and predicted phenotypes using bivariate animal models in ASReml-R version 4.1 (Gilmour et al., 2015).

In those models, we included “sampling day” as a categorical fixed effect with one level for each of the five measurement days to account for the sampling bias and effect of the person who measured the production traits on a single day. Gilthead seabream is protandrous (Loukovitis et al., 2011), and all individuals were males at harvest. Therefore, we did not include “sex” as a fixed effect in our analyses.

3.2.5 Comparing accuracy of GEBVs between genomic selection designs

For each trait, we compared the accuracy of GEBVs between GS_S, and GS_PP. In GS_S, the reference population was assumed to be slaughtered for

phenotyping and genotyped, whereas selection candidates were genotyped but not phenotyped. Using phenotypes of reference population and genotypes of reference population and selection candidates, GEBVs were estimated. In GS_PP, ungenotyped fish were assumed to be imaged before slaughtering for phenotyping, which was used to build phenotypic prediction models. Then the phenotypes of genotyped selection candidates were predicted using these models, and GEBVs were estimated using the predicted phenotypes as performance data. A visualization of GS_S and GS_PP can be found in Figure 3.1 and Figure 3.2, respectively.

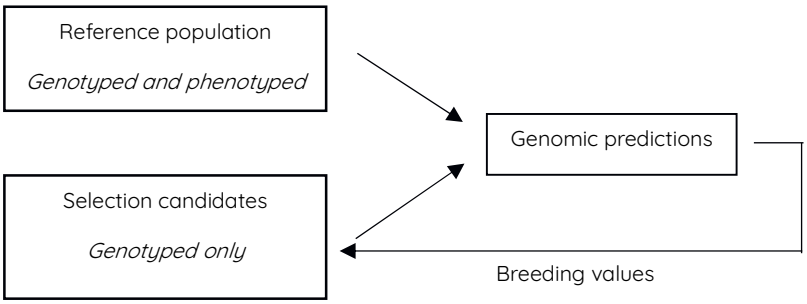

Figure 3.1. Standard genomic selection design. In this design slaughtered fish are used as reference population to predict breeding values for slaughter traits on selection candidates.

In both designs, we estimated the accuracy of GEBVs using 20-fold repeated random subsampling cross-validation. In each repeat, we split the data into two equal sized parts and treated one part as the reference population (slaughtered fish) and the other part as the selection candidates. For GS_S, the reference population was phenotyped and genotyped, and selection candidates were genotyped but not phenotyped. For GS_PP, the slaughtered fish was phenotyped but not genotyped. Phenotypes of selection candidates were predicted as described in "*prediction of phenotypes*". GEBVs for the selection candidates were

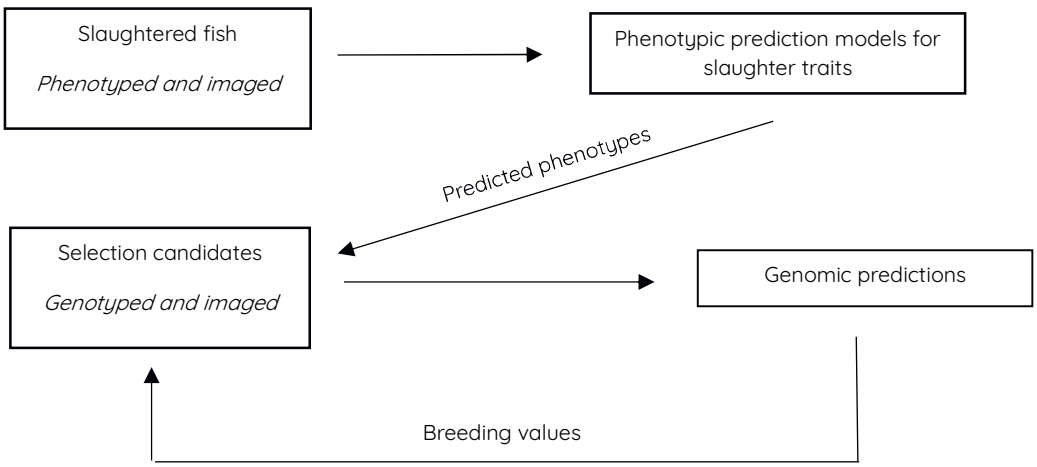

Figure 3.2. Genomic selection design with predicted phenotypes. In this scheme, slaughtered fish are not genotyped but used to build phenotypic prediction models which can be used to predict phenotypes directly on selection candidates.

estimated using a univariate animal model in ASReml 4.1. This model was $y = X\beta + Zu + e$, where y is a vector of phenotypes of the reference population in GS_S or the predicted phenotypes of selection candidates in GS_PP, β is the vector of the fixed effect of "sampling day", u is the vector of random animal additive genetic effects $\sim(0, G\sigma_a^2)$ where *G* is the GRM, σ_a^2 is the genetic variance, and e is the vector of random residual effects $\sim(0, I\sigma_e^2)$ where *I* is an identity matrix. For each repeat, we estimated the accuracy of the breeding values separately as $\sqrt{1 - (se^2/\sigma_a^2)}$ where *se* was the mean standard errors of GEBVs of the selection candidates and averaged the accuracies.

3.3 Results

The results for the phenotypic prediction models are presented in "*Prediction of phenotypes*" and the summary statistics for the predicted and observed phenotypes are presented in section "*Summary statistics*".

3.3.1 Prediction of phenotypes

The final model for each trait using all data is presented in Table 3.1. The highest accuracy was always achieved with two predictors. In Table 3.1, accuracy of the models with a single predictor is presented in addition to the final accuracies (reduced model). The highest accuracy of the phenotypic prediction was obtained for harvest weight. Weight traits were predicted with higher accuracies than ratio traits.

Table 3.1. Phenotypic prediction accuracies for final model and reduced model.

Trait	Model	Final model	Reduced model	N*
Harvest weight (g)	-75.02 + 0.55 x P_1 + 0.02 x P_2	0.98	0.97	932
Fillet weight (g)	-52.57 + 0.24 x P_1 + 0.01 x P_2	0.92	0.90	932
Fillet %	52.83 – 0.78 x P_3 + 0.10 x P_4	0.27	0.23	928
Fillet fat (%)	3.98 + 0.60 x P_5 - 0.01 x P_6	0.53	0.40	927
Viscera weight (g)	-12.48 + 0.05 x P_1 + 0.75 x P_5	0.78	0.77	928
Viscerosomatic index (%)	7.65 + 0.16 x P_5 - 0.03 x P_7	0.36	0.26	926

P_1 = Trunk volume, P_2 = Total area, P_3 = Relative head area, P_4 = Head length, P_5 = 3rd equidistant thickness, P_6 = Head area, P_7 = Tail excluded length

Reduced model is a simple regression model with the highest prediction accuracy. N* = Number of available records for both predictors in each model.

3.3.2 Summary statistics

Summary statistics for observed and predicted harvest weight, fillet weight, fillet yield, and fillet fat (%), viscera weight, and viscerosomatic index are shown in Table 3.2. The means of observed and predicted phenotypes were equal. However, the coefficient of variation (CV) of the predicted phenotypes was lower, most notably for ratio traits.

Table 3.2. Means and coefficients of variation (CV) of observed and predicted traits.

Trait	Observed		Predicted	
	Mean	CV	Mean	CV
Harvest weight (g)	412.2	17.0	411.9	16.6
Fillet weight (g)	180.0	19.5	179.9	17.9
Fillet yield (%)	43.6	7.8	43.6	2.2
Fillet fat (%)	12.7	22.1	12.7	11.7
Viscera weight (g)	29.4	23.5	29.4	18.4
Viscerosomatic index (%)	7.1	14.7	7.1	5.2

3.3.3 Genetic parameters

The genetic parameters for observed and predicted phenotypes are presented in Table 3.3. For predicted fillet weight, viscera weight, and fillet yield, the environmental variation decreased more than genetic variation compared to observed phenotypes, resulting in higher heritabilities. Genetic correlations of predicted phenotypes to observed phenotypes are also presented in Table 3.3.

Table 3.3. The genetic variation (V_A), environmental variation (V_E), and heritability (h^2) of observed and predicted phenotypes.

Trait	Observed			Predicted			Genetic correlation
	V_A	V_E	h^2 ± se	V_A	V_E	h^2 ± se	
Harvest weight (g)	2801.46	2415.48	0.54 ± 0.05	2602.71	2389.27	0.52 ± 0.05	1.0 ± 0.001
Fillet weight (g)	589.19	659.03	0.47 ± 0.05	565.58	526.65	0.52 ± 0.05	0.99 ± 0.01
Fillet %	1.02	7.49	0.12 ± 0.04	0.28	0.62	0.31 ± 0.05	0.74 ± 0.13
Fillet fat (%)	3.36	3.35	0.50 ± 0.05	0.82	1.21	0.40 ± 0.05	0.65 ± 0.07
Viscera weight (g)	23.34	27.95	0.46 ± 0.05	15.56	15.59	0.50 ± 0.05	0.79 ± 0.04
Viscerosomatic index (%)	0.48	0.58	0.45 ± 0.05	0.04	0.08	0.36 ± 0.05	0.22 ± 0.12

3.3.4 Accuracy of GEBVs in genomic selection designs GS_S and GS_PP

In GS_PP designs, phenotypic prediction models were fit in each replicate using the reference population data, which is half of the population. In those replicates, the most frequently observed predictors were the same as in Table 3.1, except for fillet fat percentage. For fillet fat percentage, the second most frequently observed predictor was *thickness by height* instead of *head area*. The mean phenotypic prediction accuracies were similar irrespective of using half of the data or all available data (Table 3.4).

The accuracy of the GEBVs in GS_S and GS_PP designs is presented in Table 3.4. For all traits, the accuracy of the GEBVs was higher in GS_PP than in GS_S, most notably for fillet yield. The standard deviations across replicates were slightly higher when half of the data was used (GS_PP).

Table 3.4. The accuracy of the GEBVs in GS_S and GS_PP designs, and phenotypic prediction accuracies in GS_PP designs (inside brackets are standard deviations).

Trait	Accuracy of the GEBVs		Phenotypic prediction accuracies in GS_PP
	GS_S*	GS_PP*	
Harvest weight (g)	0.75 (0.02)	0.81 (0.03)	0.98 (0.00)
Fillet weight (g)	0.72 (0.03)	0.81 (0.03)	0.91 (0.01)
Fillet %	0.50 (0.09)	0.72 (0.05)	0.24 (0.04)
Fillet fat (%)	0.75 (0.02)	0.76 (0.03)	0.50 (0.03)
Viscera weight (g)	0.73 (0.02)	0.80 (0.02)	0.78 (0.02)
Viscerosomatic index (%)	0.74 (0.02)	0.75 (0.04)	0.34 (0.05)

*GS_S = Standard genomic selection, GS_PP = Genomic selection with predicted phenotypes

3.4 Discussion

In this study, we introduced a novel approach using predicted phenotypes of selection candidates as performance data in genomic prediction models. Using seabream data, we demonstrated that predicting phenotypes with image analysis could significantly reduce the cost of genomic selection by eliminating genotyping costs for the reference population. In our dataset, GS_PP produced higher accuracies for GEBVs than GS_S with our approach, in which we took half of the population as selection candidates.

Using our dataset, we show that genomic selection using predicted phenotypes can accelerate genetic improvement compared to standard genomic selection. It should be noted that the GEBV accuracies in GS_S depends largely on the size of the reference population (Daetwyler et al., 2008) and the accuracy of the GEBVs improves gradually with increasing the size of the reference population (Dagnachew and Meuwissen, 2019). If enough resources are available, almost unity GEBV accuracies can be obtained by using large reference populations. However, in GS_PP design increasing the number of selection candidates has a stronger effect on the GEBV accuracies because selection candidates in GS_PP contribute performance data to the breeding value estimations. In addition, increasing the number of selection candidates also increases the selection intensity, which in combination should lead to higher selection responses in GS_PP.

Heritability is one of the determinants of the accuracy of GEBVs and higher heritabilities leads to higher accuracies (Daetwyler et al., 2008; Nielsen et al., 2009). Therefore, the higher accuracies of GEBVs in GS_PP could partially be attributed to the higher heritability of predicted phenotypes, especially for fillet yield. The predicted phenotypes are expected to be more uniform than observed phenotypes because they are built using automated morphometric measurements, which are less prone to measurement errors than manual phenotyping (Navarro et al., 2016). Manual measurement of fillet yield is known

to be especially prone to errors (Rutten et al., 2004), which may have contributed to higher heritability of predicted fillet yield. In addition, the standard error of the genetic correlation between observed and predicted fillet yield was higher compared to other traits, which may also have been caused by the difference between the uniformity of observed and predicted phenotypes.

For weight traits and fillet fat percentage, there were not large differences between the heritabilities of observed and predicted phenotypes, but the accuracies in GS_PP were still higher for these traits. In GS_S and GS_PP, the number of phenotypes available was the same and the difference was that in GS_PP, the selection candidates had their own predicted performance records, which may explain the higher accuracy of the GS_PP design. This is consistent with the findings of García-Ballesteros et al., (2022), who compared the accuracy of GEBVs for scenarios where a sib trait was measured on the sibs of the selection candidates versus measuring an indicator trait on the same number of selection candidates. In their study, the accuracy of the GEBVs for the indicator trait was higher when the heritability of the sib trait and indicator trait were equal. In our study, the predicted phenotypes are correlated indicator traits for observed phenotypes.

The GS_PP design reduces the genotyping effort considerably by removing the need to genotype a reference population. The phenotyping effort may also reduce by reducing the number of slaughtered fish; however, phenotypic prediction accuracies may deteriorate. In our analyses, reducing the number of slaughtered fish from 900 (Table 3.1) to 450 (Table 3.4) did not lead to a noteworthy deterioration in phenotypic prediction accuracies. However, further reduction of the number of slaughtered fish may affect the accuracy unfavorably. Therefore, attention should be paid to slaughtering an adequate number of fish to obtain optimum phenotypic prediction accuracies. Phenotyping effort can further be reduced by reusing the slaughtered fish in future generations, which eliminates the need to slaughter fish in every generation. The

advantage of this is that considerably less labor is required in the breeding program. Phenotypic predictions might be relatively robust once established, although it cannot be excluded that morphometric dimensions change because of selection. In common sole, Blonk et al. (2010) showed that selection will change the shape of fish. A similar trend was observed for Nile tilapia (Trọng et al., 2013). In this study, the selection candidates and slaughtered fish originated from the same environment and were closely related. However, in many fish breeding programs, the selection candidates are generally kept in a different environment than slaughtered fish. Genotype by environment interaction can occur between different production systems (Khaw et al., 2012; Sang et al., 2020), which can affect the shape of the fish and proportions of the body between environments. For example, Trọng et al., (2013) has shown that there is substantial GxE for shape traits in Nile tilapia between different production systems. This may result in lower accuracy in phenotypic predictions of selection candidates, which should be tested and validated in further studies.

Predicting phenotypes with digital images can be advantageous in pedigree-based breeding programs when sibs are slaughtered to measure sib traits. These sib populations can be used as training data for phenotypic prediction models, which can then be used to predict own performance of selection candidates. This design allows exploiting Mendelian variance, which should result in increased selection responses.

3.4.1 Phenotypic predictions

The breeding goal in fish breeding programs may include only growth, which is strongly correlated to harvest weight (>0.99) (Gulzari et al., 2022), or a combination of growth and other traits. If growth is the only trait in the breeding goal, capturing images of the fish to predict harvest weight seems like an unneeded effort. However, imaging the fish could still be relevant in this case. If other traits are added to the breeding goal in later generations, the images of

the ancestors can be used to predict ancestral phenotypes, which may increase the accuracy of EBVs in offspring.

Another advantage of using predicted rather than observed weights is that predictions are far less prone to bias and error. Taking accurate weights on live fish is notoriously difficult, and stressful to the animals. High throughput imaging should lead to higher precision, which is shown in Table 3.2 where CV of observed phenotypes was in all cases higher than the CV of predicted phenotypes.

To our knowledge, the phenotypic prediction accuracy for estimated fillet traits in seabream has not been reported before using morphometric measurements in multiple regression models. However, Sang et al. (2009) investigated the prediction accuracy of multiple regression models for river catfish (*Pangasionodon hypopthalmus*) with manual morphometric measurements as predictors. Similar to our study, *volume* of the fish was the most important predictor in their study for fillet traits. The phenotypic prediction accuracy for fillet weight in this study agrees well with the value of 0.93 reported by Sang et al. (2009). However, they reported a much higher prediction accuracy for fillet yield (0.86) than in this study. An explanation for this difference might be that the elongated shape of catfish enables more accurate predictions from morphometric measurements, especially volume. High prediction accuracy found in this study for fillet weight (0.92) is slightly lower than the value reported by Navarro et al., (2016) (0.97), who used *fillet area* in gilthead seabream. This indicates that the predictive power of the morphometric measurements may differ between unrelated populations of seabream. *Relative head area* was the morphometric measurement with the highest predictive power for fillet yield. This result is consistent with the findings in common carp (*Cyprinus carpio*) (Prchal et al., 2020, 2018).

3.5 Conclusions

Our findings show that image analysis can reduce the genotyping costs of a GS breeding program and simultaneously improve the accuracy of the GEBVs. The

accuracy of the GEBVs was higher in GS_PP designs for all production traits, in which slaughtered fish contributed images and observed phenotypes, and genetic analyses were performed with predicted phenotypes of selection candidates.

Acknowledgement

Authors would like to thank Cudomar for their participation in the study and their staff for providing necessary help and facilities for raising the population and the experimental data collection.

Funding

This research received funding from the European Commission Horizon 2020 (H2020) Framework Program through grant agreement no. 727315 for the MedAID project (Mediterranean Aquaculture Integrated Development) and no. 818367 for the AquaIMPACT project.

Chapter 4:

Optimizing Phenotyping Effort in Two-Stage Genomic Selection Schemes for Difficult-To-Measure Traits in Fish Breeding Programs

Benan Gulzari[a], Chantal Roozeboom[a], Mark Camara[b],

Hans Komen[a], John W.M. Bastiaansen[a]

[a]Department of Animal Breeding and Genomics, Wageningen University and Research, Droevendaalsesteeg 1, 6700 AH, Wageningen (The Netherlands)

[b]Wageningen Livestock Research, Wageningen University and Research, Droevendaalsesteeg 1, 6700 AH, Wageningen (The Netherlands)

Abstract

In pedigree-based breeding programs, selection candidates typically do not have offspring and therefore get family-average breeding values for traits that can only be phenotyped on their siblings. Genomic selection (GS) overcomes this drawback by capturing Mendelian sampling effect within families. Despite this advantage, GS is not widely used in fish breeding programs. The performance of GS programs depends largely on genotyping and phenotyping strategies. Much research has focused on genotyping strategies; however, phenotyping strategies are also influential on the performance of GS. Therefore, we aimed to test the effect of phenotyping effort on the selection response in a difficult-to-measure trait (DIF) and harvest weight (HW) in a genomic selection program with fixed genotyping effort. We simulated a two-stage selection scheme in which all candidates were genotyped, and varying proportions were preselected on phenotypic harvest weight and unselected candidates were slaughtered to obtain data on DIF. Final selection was on an index with equal weights on genomic breeding values for HW and DIF. The DIF had either a positive (0.4) or negative (-0.1) genetic correlation to HW. We simulated traits with additive gene effects for 1,000 QTL for HW and an additional 1,000 QTL for DIF. In the first generation, the maximum response in selection index was obtained with a preselected proportion of 40% when $r_g = 0.4$ and 50% when $r_g = -0.1$. For both levels of genetic correlations, the responses in selection index with all other preselected proportions were close to the maximum response. In the first generation, the maximum selection response in DIF was obtained with a preselected proportion of 40% independent of genetic correlation. Between 20% and 60% preselection the response in DIF was close to the maximum response. At 40% preselected proportion, the response in HW was close to its minimum. The maximum response in HW was obtained at 90% preselection. The h^2 of DIF reduced mostly linearly at a rate of 0.9% per generation when $r_g = 0.4$ and 1% when $r_g = -0.1$. The h^2 of HW reduced mostly linearly at a rate of 2.5% per

generation independent of r_g. In conclusion, results show that the phenotyping effort for DIF can be adjusted while maintaining a close-to-maximum response in selection index and DIF.

4.1 Introduction

Aquaculture has increasingly been contributing to the production of animal protein (FAO, 2021), and to maintain this increasing trend, high yielding genetic material is required. Genetic improvement in aquaculture is widely achieved with pedigree-based selection methods (Bentsen et al., 2017; Camara and Symonds, 2014; Chavanne et al., 2016). However, pedigree-based methods have drawbacks. For so called difficult-to-measure traits (DIF), candidates are assigned family-average breeding values based on phenotypes of the siblings. Because pedigree relationships cannot capture the genetic differences between siblings (Mendelian sampling effects), the accuracy of the EBVs can be low depending on the heritability and the number of siblings.

Genomic selection is a relatively recent technology, in which the breeding values are estimated either by summing the estimates of genome-wide genetic marker effects or using relatedness of the animals estimated using marker information in genomic best linear unbiased prediction (GBLUP) (Meuwissen et al., 2001). Genomic selection captures the Mendelian sampling effects meaning that the candidates within families can be ranked without their own phenotypes for DIF. Despite the advantages, genomic selection is not widespread in fish breeding programs mainly because of high genotyping costs (Boudry et al., 2021). Much research has focused on strategies to lower these costs, including using low density SNP panels (Kriaridou et al., 2020), SNP imputation (Bolormaa et al., 2015), and reducing sib testing (Sonesson and Meuwissen, 2009).

Difficult-to-measure traits are included in some fish breeding programs such as fillet yield (Vandeputte et al., 2019), fillet pigmentation (Ødegård et al., 2014) and fatty acid composition (Horn et al., 2020). These traits may be favorably correlated with production traits such as harvest weight, for example fillet yield in gilthead seabream (Gulzari et al., 2022), or unfavorably correlated, for example resistance to Piscirickettsia salmonis in coho salmon (Yáñez et al., 2016). In both cases, accuracy of GEBVs for DIF increases with more animals in the reference

population (García-Ballesteros et al., 2022); however, genotyping large reference populations is costly and phenotyping may be laborious. Genotyping effort is typically limited, meaning that the division of genotyped animals into reference population and selection candidates should be optimized (Chu et al., 2018).

In this study, we investigated the optimal division of animals into selection candidates and reference population to obtain the desired selection responses for selection index, harvest weight (HW), and DIF under constrained genotyping efforts. To determine this optimal division, we simulated a two-stage selection scheme in which all candidates were genotyped, and varying proportions were preselected on phenotypic harvest weight. Unselected candidates were slaughtered to obtain data on DIF, which was either positively or negatively correlated to harvest weight.

4.2 Materials & Methods

We simulated breeding programs with the Modular Breeding Program Simulator (MoBPS) package in R software (Pook et al., 2020; R Core Team, 2020). We simulated fish with breeding values and phenotypes for HW and a DIF. Two-stage selection was performed. The first stage is phenotypic selection on HW. In the second stage, the selection is on an index of DIF and HW using genomic estimated breeding values (GEBVs).

4.2.1 Simulation of base populations

Initially, we simulated 100 male and 100 female diploid individuals that had 24 equal sized chromosomes with a total genome size of 15 Morgan (M) and total physical size of 500 mega bases (Mb). This approximates the genome structure of some commonly farmed fish species, such as gilthead seabream (*Sparus aurata*), European seabass (*Dicentrarchus labrax*), and turbot (*Psetta maxima*) (Aslam et al., 2020; Fernández et al., 2021; Tine et al., 2014). We simulated 1600 equidistant biallelic loci per chromosome that were in linkage equilibrium. Genotypes for each locus had initial allele frequencies of 0.5. To create linkage disequilibrium (LD) in relatively few generations, we simulated 500 generations

where we randomly selected 50 male and 50 female parents to create 100 male and 100 female offspring. Each fish was mated once and the number of offspring per mating was not restricted. We set the mutation rate to 10^{-3} per locus across all generations. Mutations occurred randomly and changed an allele of one of the existing variants. At the end of 500 generations, the heterozygosity stabilized at 0.25. The baseline simulation process was repeated 100 times. At the last generation of those 100 replicates, on average there were 30,000 segregating loci. The physical distance between adjacent segregating loci on each chromosome was on average 0.017 Mb that had an LD of 0.15 measured as r^2. From the last generation, we randomly selected 50 male and 100 female parents and mated them randomly in a 1:2 mating design to produce 30 offspring per female a base population of 1,500 males and 1,500 females as the first generation of selection candidates. Genomic selection scenarios were simulated with the mutation rate set to zero.

4.2.2 Simulation of traits

In the base population, we simulated HW and a DIF for each candidate with additive gene effects only. HW was controlled by 1,000 randomly selected SNPs as quantitative trait loci (QTL), with an initial heritability (h^2) of 0.3. The same QTL also controlled DIF together with 1,000 additional QTL, with an initial h^2 of 0.1. The effects of the QTL were drawn from a normal distribution with $N(0,1)$ separately for two traits and then scaled to obtain the desired genetic variances and covariance. True breeding values (TBV) of the candidates were calculated by summing additive genetic effects of the QTLs. The genetic correlation of DIF to HW was either positive (0.4) or negative (-0.1), and the phenotypic correlation between the traits was 75% of the genetic correlation. We calculated the correlation between the environmental values of the traits as $\rho_{E_{t_1},E_{t_2}} = \frac{\rho_{P_{t_1},P_{t_2}} - (\sqrt{h^2_{t_1} h^2_{t_2}}\, \rho_{TBV_{t_1},TBV_{t_2}})}{\sqrt{(1-h^2_{t_1})(1-h^2_{t_2})}}$, where $\rho_{P_{t_1},P_{t_2}}$ and $\rho_{TBV_{t_1},TBV_{t_2}}$ are the phenotypic and genetic correlations between the traits HW and DIF. The correlations between

genetic and environmental effects were zero. Residual variances (σ_e^2) for the traits were calculated as $\sigma_e^2 = \sigma_p^2 - \sigma_a^2$, where σ_p^2 is the phenotypic variance and σ_a^2 is the additive genetic variance. σ_p^2 was set to one for each trait. We kept the residual variances constant across generations. Genetic variances, and therefore h^2 changed over time due to effects of selection and drift.

4.2.3 Selection strategies

We simulated a two-stage genomic selection program. The first stage applied phenotypic selection on HW in which all fish were available as selection candidates. Only the fish selected in the first stage (preselected) were available for selection in the second (final) stage. We varied the selected proportion in the first stage from 10% to 90% with increments of 10% and adjusted the selected proportion in the second stage to finally select 50 males and 100 females. The fish that were not selected in the first stage were assumed to be slaughtered to measure DIF, thus serving as a reference population. All 3,000 candidates were genotyped, and genomic breeding values of the preselected fish were estimated for both traits. We computed the genomic relationship matrix GRM as $G = \frac{ZZ'}{2\sum_{p_i}(1-p_i)}$ (VanRaden, 2008), Z is a matrix that contains marker genotypes for all loci, which is corrected for the allele frequency per locus and, p_i is the minor allele frequency. We estimated the breeding values and variance components using univariate animal models. The univariate animal model is $y = \mu + Zu + e$, where y is a vector of phenotypes for HW or DIF, μ is the trait mean, u is the vector of random animal additive genetic effects $\sim(0,\ G\sigma_a^2)$, G is the genomic relationship matrix and σ_a^2 is the additive genetic variance of the trait, and e is the vector of random residual effects $\sim(0,\ I\sigma_e^2)$, I is an identity matrix, σ_e^2 is the residual variance of the trait. Z is a design matrix that relates observations to the additive genetic effect of animals. The expectation of $cov(u, e)$ is zero. We calculated the heritability of each trait as $h^2 = \frac{\sigma_a^2}{\sigma_a^2+\sigma_e^2}$ and the accuracy of GEBVs as the Pearson correlation with the TBVs of preselected fish as $r = \frac{\sum(TBV_i-\overline{TBV})(GEBV_i-\overline{GEBV})}{\sqrt{\sum(TBV_i-\overline{TBV})^2\sum(GEBV_i-\overline{GEBV})^2}}$.

From the preselected fish, we selected the 50 males and 100 females using a selection index in which both traits had equal weights. Selected fish were randomly mated in the 1:2 mating design while avoiding the mating of full and half siblings to create the next generation of 1,500 male and 1,500 female selection candidates. The number of offspring per mating was equal. We performed selection for 10 generations. We used each of the 100 base populations twice with different randomly assigned QTL for each trait to produce 200 replicates for every preselection scenario.

We calculated the true response to selection in genetic standard deviations for each trait as $R = \frac{\overline{TBV_{n+1}} - \overline{TBV_n}}{\sigma_{TBV_n}}$, $R$is true response to selection, $\overline{TBV_{n+1}}$ is the mean true breeding value of fish in generation $n+1$, $\overline{TBV_n}$ is the mean true breeding value of fish in generation n, and σ_{TBV_n} is the standard deviation of TBVs of fish in generation n. We estimated the rate of inbreeding in each generation as $\Delta F_n = (\overline{F_{n+1}} - \overline{F_n})/(1 - \overline{F_n})$ using pedigree-based inbreeding coefficients.

4.2.4 Evaluation of results

For all scenarios the selection responses of index, DIF, and HW, accuracy of GEBVs, heritabilities, and inbreeding rate were averaged in every generation across the 200 replicates. We also averaged the heritabilities and genetic correlations across 200 replicates to evaluate the changes. We calculated the rate of change in heritabilities and accuracies across 200 replicates using linear regression models where the number of generations was the predictor variable.

4.3 Results

4.3.1 Response to selection

We evaluated the selection results after one, five, and ten generations of selection.

In the first generation, the maximum response in selection index was obtained with a preselected proportion of 40% when $r_g = 0.4$ (2.57 $\sigma_{TBV_{index}}$) and 50% when

$r_g = -0.1$ (1.90 $\sigma_{TBV_{index}}$) (Figure 4.1). For both levels of genetic correlations, the responses obtained with most other preselected proportions were close to the maximum response. The exceptions were for preselected proportion of 10% (2.32 $\sigma_{TBV_{index}}$) for $r_g = 0.4$ and, 90% (1.65 $\sigma_{TBV_{index}}$) for $r_g = -0.1$. In generations five and ten, we see less response, but with the same pattern of the maximum response at an intermediate selected proportion and not much reduction with more extreme proportions.

In the first generation, the maximum selection response in DIF was obtained with a preselected proportion of 40%. This optimal proportion was independent of the initial genetic correlation. The level of response was higher when $r_g = 0.4$ (1.1 $\sigma_{TBV_{DIF}}$) than when $r_g = -0.1$ (0.4 $\sigma_{TBV_{DIF}}$) (Figure 4.2). For both levels of genetic correlations, the responses obtained with preselected proportions between 20% and 60% were close to the maximum response. The response in DIF decreased considerably as the preselected proportion reduced below 20% or increased above 60%. The preselected proportion of 90% yielded the least response of 0.8 $\sigma_{TBV_{DIF}}$ when $r_g = 0.4$ and 0.03 $\sigma_{TBV_{DIF}}$ when $r_g = -0.1$. In generations five and ten, we see the same pattern of maximum responses between preselected proportions of 20 and 60%. However, the magnitude of the responses was lower in later generations.

In all generations, the maximum selection response in HW was obtained with a preselected proportion of 90% (Figure 4.3). The level of response was slightly higher when $r_g = 0.4$ than when $r_g = -0.1$. The difference was small with a response of 1.643 $\sigma_{TBV_{HW}}$ for $r_g = 0.4$ and 1.636 $\sigma_{TBV_{HW}}$ for $r_g = -0.1$ in the first generation. For all generations, the response at preselected proportions between 10 and 50% were about 10% lower than at a preselected proportion of 90%. At 40% preselected proportion, where the highest response in DIF was observed, the response in HW was close to its minimum.

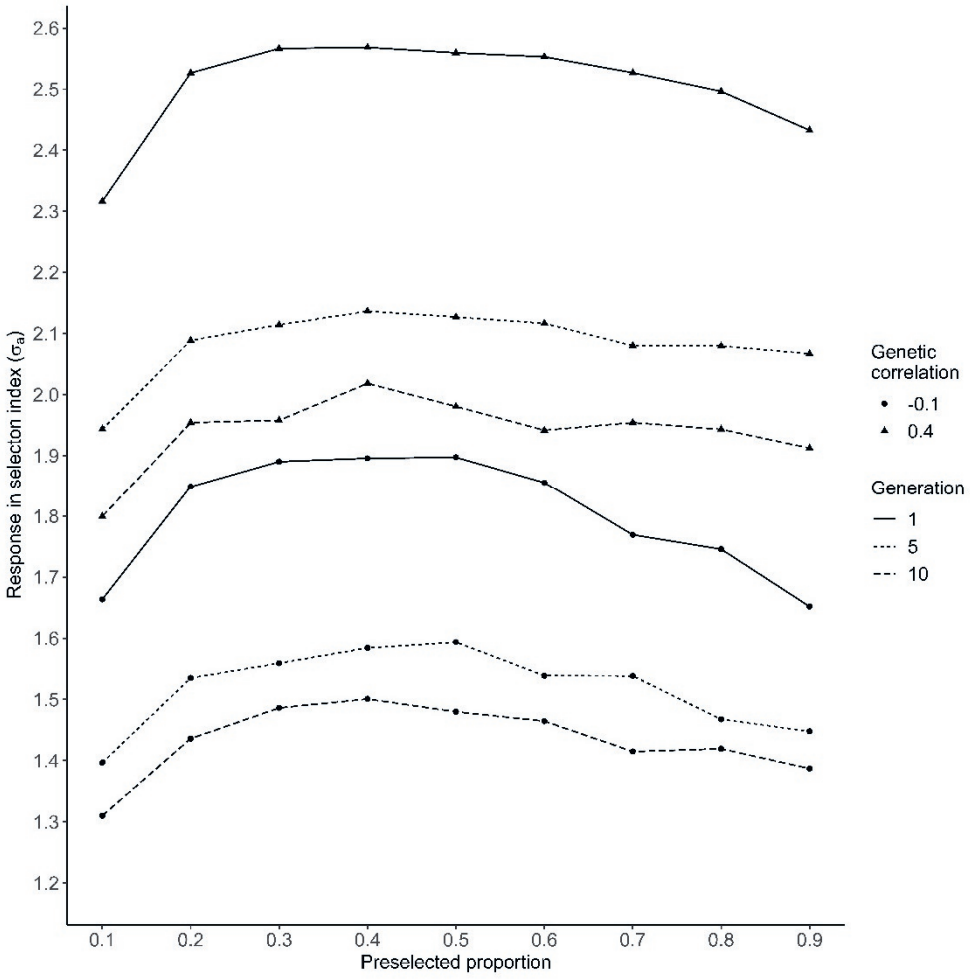

Figure 4.1. Selection response in selection index for different preselected proportions after 1, 5, and 10 generations of selection.

4.3.2 Genetic parameters and accuracy

The h^2 of DIF and HW reduced over ten generations of selection. After a larger initial reduction in the first generation, the h^2 reduced mostly linearly. The reduction was slower for DIF at a rate of 0.9% per generation when $r_g = 0.4$ (Figure 4.4) and 1% when $r_g = -0.1$ from generation 2 to 10. For HW, the reduction was faster with a rate of 2.5% per generation independent of r_g (Figure 4.5).

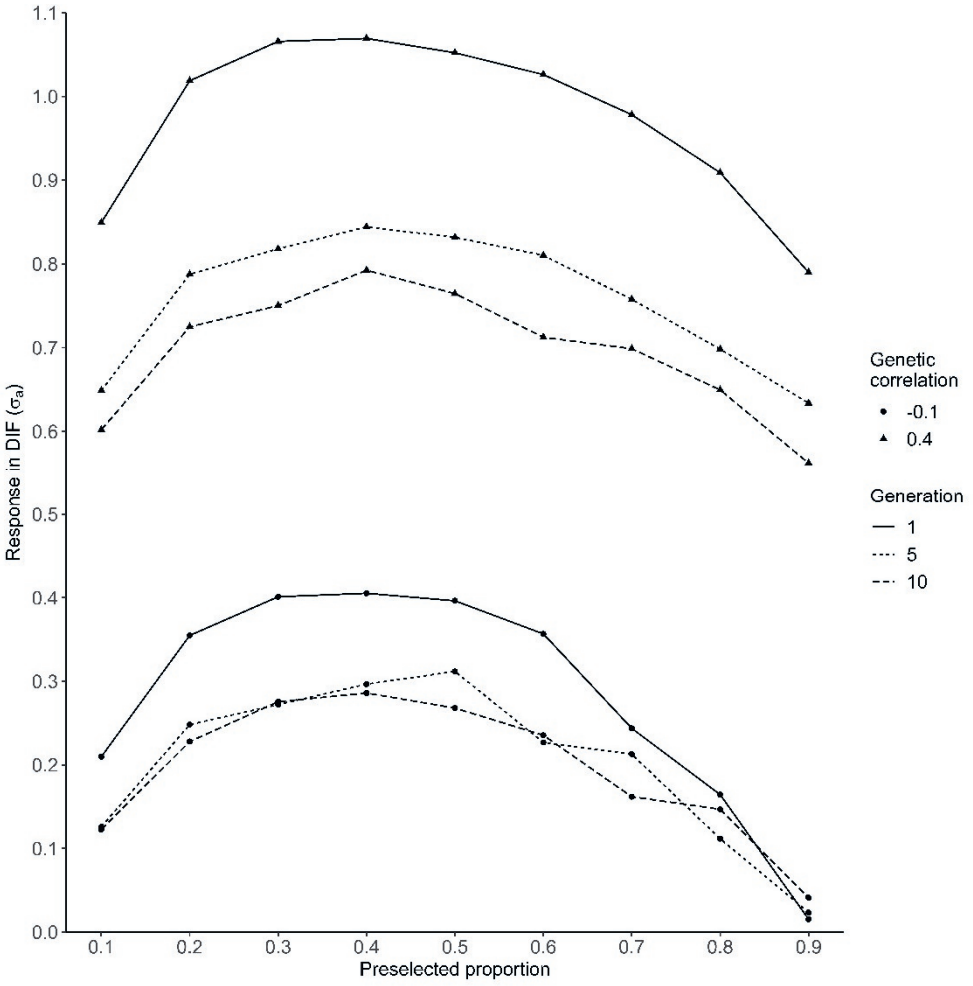

Figure 4.2. Selection response in DIF for different preselected proportions after 1, 5, and 10 generations of selection.

Accuracy of GEBVs for DIF increased when preselected proportion decreased (Figure 4.6). Opposite to what we see for DIF, accuracy of GEBVs for HW decreased when preselected proportion decreased (Figure 4.7), but the differences in accuracy between preselected proportions were small. For DIF, the number of phenotypes increased with higher preselected proportions, but for HW

the number of fish genotyped and phenotyped was equal for all levels of preselection. The accuracy of GEBVs was highest in the first generation for most

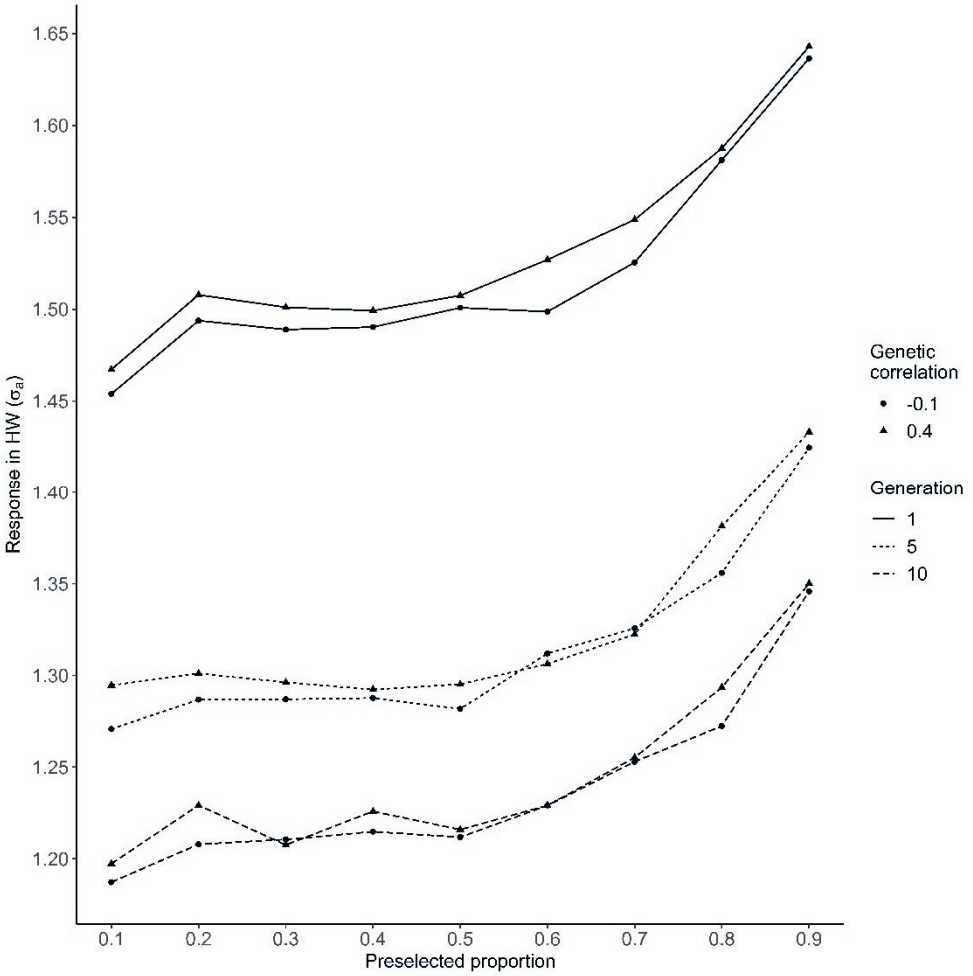

Figure 4.3. Selection response in HW for different preselected proportions after 1, 5, and 10 generations of selection.

proportions but did not change much over time for DIF; however, reduced for HW. Accuracies for DIF in the tenth generation were between 0.56 for 10% preselected and 0.30 for 90% preselected when $r_g = 0.4$. Accuracies for HW in the tenth generation were between 0.65 for 10% preselected and 0.71 for 90%

preselected when $r_g = 0.4$ with a rate of reduction of ~6.8% from generation 2 to 10 for all preselected proportions.

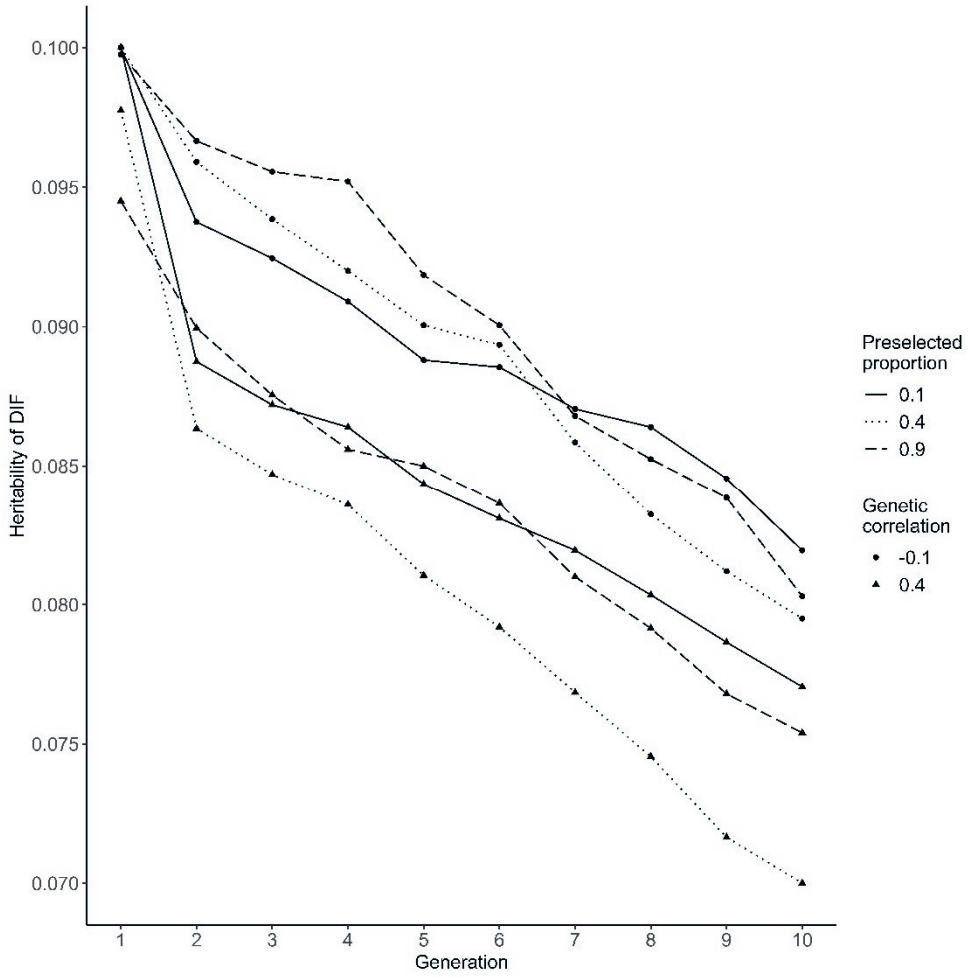

Figure 4.4. Heritability of DIF from generation 1 to 10 for preselected proportions of 0.1, 0.4, and 0.9.

4.3.3 Genetic correlation, linkage disequilibrium, inbreeding

The genetic correlation between DIF and HW changed in a negative direction, mostly linearly with selection. Like for genetic variance and heritability, the

biggest reduction was seen in the first generation of selection. The biggest effect on the genetic correlation was seen with 50% preselection where the r_g of 0.4 reduced to 0.23, and the r_g of -0.1 reduced to -0.21 after 10 generations of selection. Linkage disequilibrium increased every generation and reached a value of r^2 = 0.30 between adjacent SNPs in the tenth generation. The increase in r^2 was the same for all levels of preselected proportion.

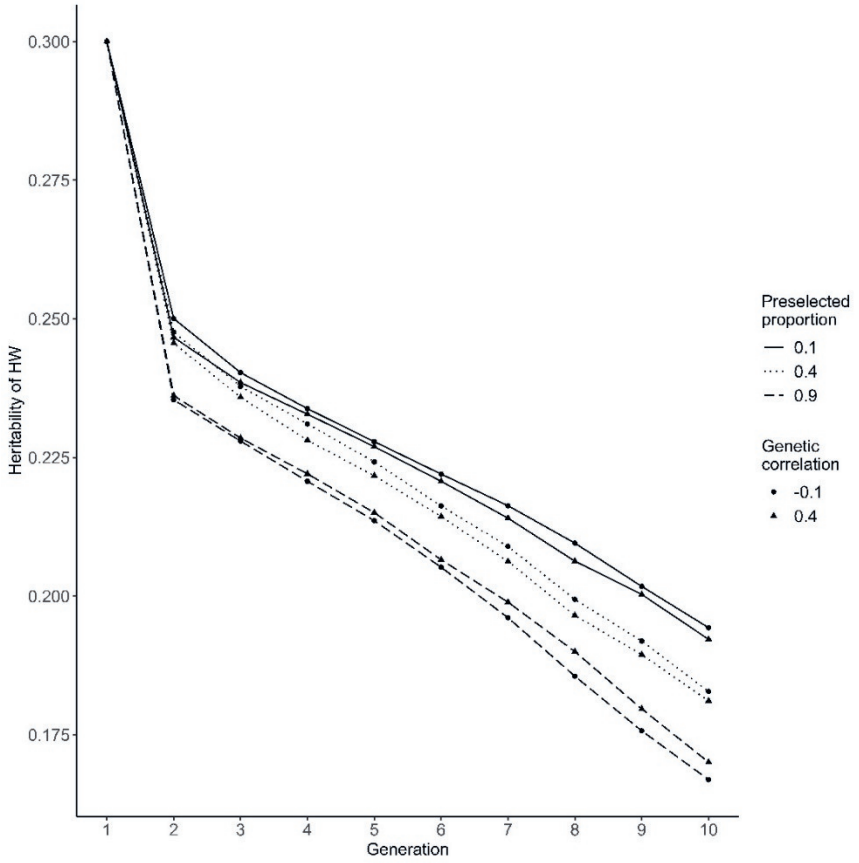

Figure 4.5. Heritability of HW from generation 1 to 10 for preselected proportions of 0.1, 0.4, and 0.9.

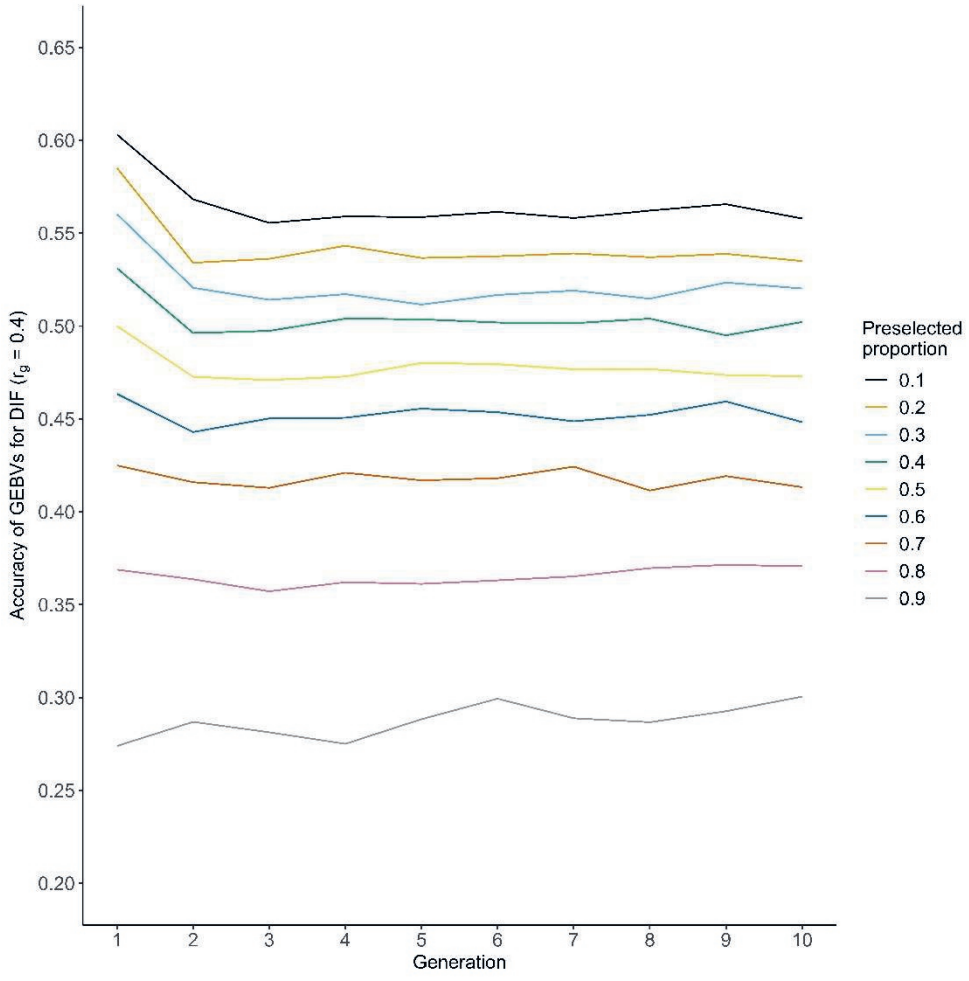

Figure 4.6. Accuracy of GEBVs for DIF from generation 1 to 10 for different preselected proportions when $r_g = 0.4$.

The rate of inbreeding (Δ_F) was independent of the genetic correlation between traits (Figure 4.8). Δ_F increased to levels around 1.5% and 2.2% in generations four and nine, respectively, for preselection of 40%.

4.4 Discussion

For a two-stage genomic selection program with a fixed genotyping effort, we investigated the selection response in a difficult-to-measure trait and harvest

weight using computer simulations. A two-stage selection scheme was simulated in which fish were first selected on phenotypic harvest weight with subsequent slaughtering of unselected fish. The proportions of selected and unselected fish varied between 10% and 90%. All fish, selected and slaughtered, were genotyped.

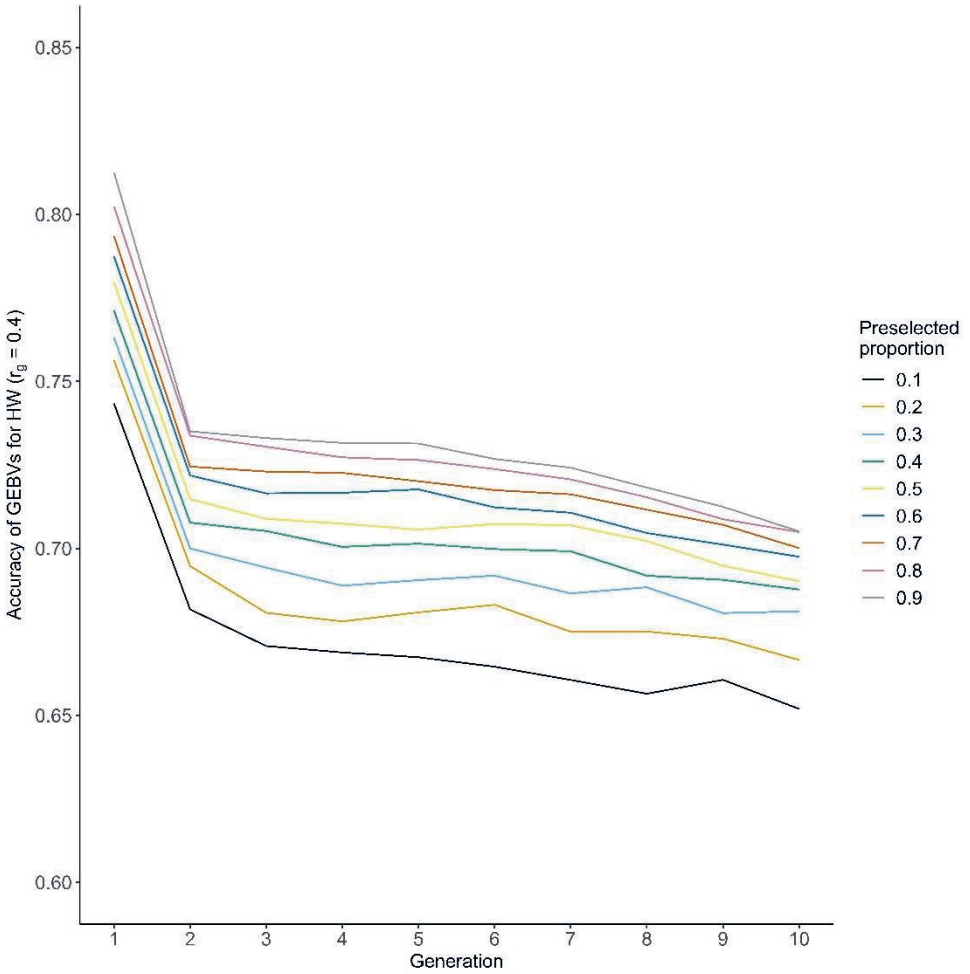

Figure 4.7. Accuracy of GEBVs for HW from generation 1 to 10 for different preselected proportions when $r_g = 0.4$.

Therefore, the number of genotyped fish was the same for each preselected proportion and all scenarios can be compared at equal genotyping efforts.

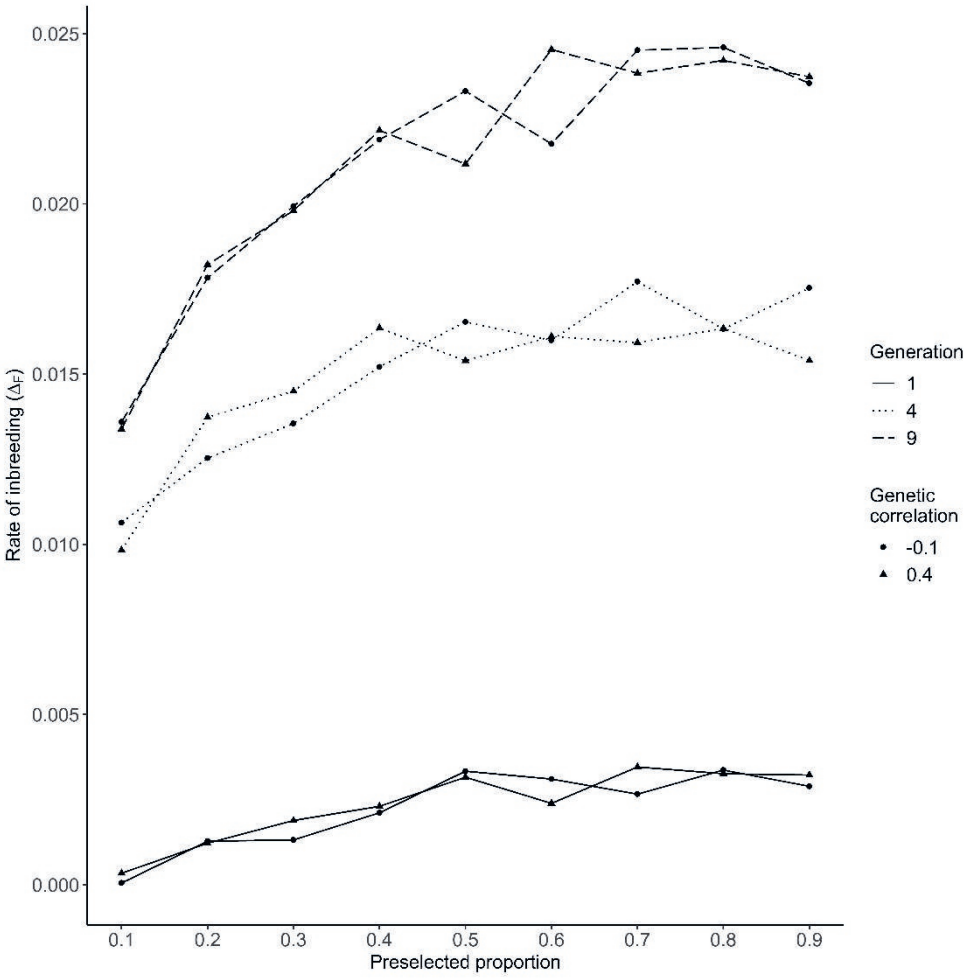

Figure 4.8. The rate of inbreeding for different preselected proportions

In this study the response in selection index is very similar for selected proportions starting at 20% up to 80-90%. The proportion with maximum response (40%) is the same for the selection index and DIF. While a breeding program typically aims to maximize the response in the selection index it may be desirable to move the preselected proportion away from 40% for several reasons. Increasing the preselected proportion will reduce the number of fish that need to be phenotyped for the difficult-to-measure trait. A difficult-to-measure trait may require slaughter or another time-consuming process and/or may

require expensive assays to obtain the phenotype data. This phenotyping cost and effort is cut in half, from 1800 fish to 900 fish in our scenarios, when the preselected proportion is increased from 40% to 70%. Decreasing the preselected proportion on the other hand, will increase the phenotyping cost and effort but this will yield a higher accuracy of the GEBV for DIF. Especially when the reference population for DIF is accumulating over multiple generations a smaller selected proportion will result in more fish being added to the reference in each generation and this will result in higher accuracy GEBV for DIF on the selected candidates. From Figure 4.1, it is clear that these choices can be made without much effect on the response of the selection index in the current generation.

While the response in selection index was mostly the same for selected proportions from 20% to 80-90%, the underlying response of DIF and HW were affected more by changing the proportion that is preselected. A higher response in DIF is offset by a lower response in harvest weight, or vice-versa, resulting in a more or less equal response of the index. The breeding goal traits are weighed in the selection index either to reflect their desired gains or by their economic values that reflect the additional income generated by improvement in each of these traits (Janssen et al., 2017). With accurate economic values the response in selection index should be maximized. However, often the information to calculate the weights is incomplete or unreliable or breeders make a different choice for their desired gains. Harvest weight is the most valuable trait when fish farms are paid on total biomass output. However, a DIF such as fillet yield will become valuable once farms get paid on the amount of fillet. The majority of fish breeding programs aim to improve harvest weight (Chavanne et al., 2016), because a high total biomass is the most desired output of farms. Total fillet output is becoming increasingly important in farming of certain fish species, such as Nile tilapia (Nguyen et al., 2010) and river catfish (Sang et al., 2009). If the most desired output of farms is a high amount of fillet, then improving fillet yield (DIF) will generate more value. A desired gains approach can be taken when it is

expected that economic values will change in the future, while exact projections are not available. In this study, we applied equal weights on HW and DIF, and the selection response in the index was similar for all levels of preselection. With index weights that put more emphasis on DIF, the response in the index will be more sensitive to the accuracy of GEBVs for DIF. In that case, lowering preselected proportion and therefore more accurate GEBV for DIF, will increase the response of the index.

The selection scheme started with 50 full-sib families and 3,000 candidates to reflect designs of breeding programs for some commonly farmed fish species. The number of full-sib families simulated in this study is the same as in Fernández et al. (2021) and, the number of candidates in between the 10,000 used by Sonesson, (2005) and Sonesson et al., (2011), and 1,500 used by Fernández et al., (2021). Selection intensity and responses will increase when starting with more selection candidates. However, we expect that the optimum preselected proportion will still show an intermediate value. With a larger number of genotyped fish, the differences in selection responses for HW and DIF may reduce because accuracies become high for both traits at all the selected proportions.

In this study, we applied mass selection on HW in the first stage and genomic selection on HW and DIF in the second stage of selection. Both breeding goal traits were normally distributed. The heritability of HW (0.3) is an intermediate value of what is reported in literature for different fish species (Garcia-Celdran et al., 2015; Guan et al., 2016; Navarro et al., 2009). DIF represents a trait that cannot be measured on selection candidates. Examples are feed efficiency, fillet yield, fatty acid composition of fillet, and product quality traits with heritability estimates ranging from 0.03 to 0.48 (de Verdal et al., 2018; Horn et al., 2022, 2018; Navarro et al., 2009; Quinton et al., 2007). In this study, we chose an intermediate value of 0.10 for the heritability of DIF. The genetic correlation between the examples of DIF and HW varies strongly. A positive genetic correlation of 0.74

was reported between HW and fillet yield (Rutten et al., 2005). A positive genetic correlation of 0.68 was reported between HW and survival (Hamzah et al., 2017), and a negative genetic correlation of -0.50 was reported for HW and *P. salmonis* resistance (Yáñez et al., 2016). Lower positive genetic correlations were reported with protein retention (0.20) (Kause et al., 2016) and muscle fat (0.13) (Horn et al., 2022), and a negative genetic correlation with cardiosomatic index (-0.11) (Gulzari et al., 2022). We chose a moderate positive (0.4) and a low negative (-0.1) values to study correlations that are small to moderate and with opposite signs.

In our study, we updated the marker effects using the data of the latest generation exclusively. Using data from earlier generations to increase the reference population size may increase the accuracy of the GEBVs (Sonesson and Meuwissen, 2009); however, will not revoke the genotyping constraints. If data from earlier generations are included in the analyses, a certain accuracy of GEBVs may be achieved with fewer slaughtered fish in the current generation. In such scenarios, more genotyping budget could be allocated to selection candidates, which will yield a higher selection intensity. However, the value of data from earlier generations may diminish with selection (Nielsen et al., 2009; Sonesson and Meuwissen, 2009).

After a drop in the first generations due to Bulmer effect, h^2 reduced mostly linearly for both traits with a steeper decline for HW compared to DIF. Because all candidates had data on HW, the selection pressure on HW depended mostly on the preselected proportion. A higher preselected proportion meant more accurate selection on HW since more candidates pass to the second stage where selection was on GEBVs. More accurate selection on HW causes a bigger reduction in h^2 due to stronger loss of genetic variation. For DIF, the selection differential was bigger with a positive genetic correlation than with a negative genetic correlation and this results in a bigger reduction in genetic variance and h^2. The accuracy of GEBVs depends on h^2 of the traits and effective population size (Ne) (Daetwyler et al., 2010). For both traits, the h^2 reduced over time but the

accuracy of GEBVs was mostly constant with some random fluctuations. In our simulation, the reduction in h^2 is offset by the reduction in Ne, which is seen from the r^2 between adjacent SNPs that increases from 0.15 to 0.30 after 10 generations of selection.

With the proposed two-stage selection scheme, the rate of inbreeding increased considerably above 1% after a few generations. We did not apply any measures to limit the rate of inbreeding such as reducing the number of candidates selected from a single family or applying optimum contribution selection (Sonesson and Meuwissen, 2000). In actual breeding programs, there should be a constraint on the inbreeding rate. If a constraint is applied, the loss of genetic variance and reduction of heritabilities will be slower and other parameters such as the accuracy of GEBVs and rate of genetic gain will be affected. We do not expect that a constraint on inbreeding will affect the optimum preselected proportions because this optimum did not change from generation 1 to 10 while inbreeding rate was increasing and levels of genetic variance and heritability were reducing.

We assume preselection on HW and including this same trait in the selection index. Preselection at an early age is also possible and this might have a practical implication with, for instance, disease challenge tests, which are typically applied at an early age. If the time gap between the two stages of selection increases, the efficiency of the first stage selection might diminish because the genetic correlation of the early body weight with the index trait HW is reduced. A big advantage of early preselection is that only the preselected fish must be raised to harvest weight.

4.5 Conclusions

In our simulations of two-stage selection, the optimum prelected proportion was 40% to yield the highest selection response in both the selection index and the difficult-to-measure trait. The response in selection index was very similar for a wide range of preselected proportions. This means that the phenotyping effort

for DIF might be reduced to save costs, or increased to quickly build a larger reference population, without a significant loss in the selection response for index. Subsequent generations of selection and increasing levels of inbreeding did not affect the optimum preselected proportions. Therefore, the phenotyping effort for DIF can be adjusted while maintaining a close-to-maximum response in selection index and DIF.

Funding

This research received funding from the European Commission Horizon 2020 (H2020) Framework Program through grant agreement no. 818367 for the AquaIMPACT project.

Chapter 5:

Genetic Parameters and Genotype by Environment Interaction for Production Traits and Organ Weights of Gilthead Seabream (*Sparus aurata*) Reared in Sea Cages

Benan Gulzari[a], Hans Komen[a], Varun Raj Nammula[a,b], John W.M. Bastiaansen[a]

[a]Department of Animal Breeding and Genomics, Wageningen University and Research, Droevendaalsesteeg 1, 6708 PB, Wageningen (The Netherlands)
[b]Department of Integrated Plant and Animal Breeding, University of Göttingen, Wilhelmsplatz 1, 37073 Göttingen (Germany)

Aquaculture 548, 737555 (2022)

Abstract

Gilthead seabream is a key fish species for farming in the Mediterranean region and is farmed in a large geographical area under various production circumstances. However, more than 80% of the genetically improved fingerlings originate from a single country, Greece, which poses a potential risk for genotype by environment interaction (GxE). Therefore, the objective of this study was to quantify GxE for several traits of gilthead seabream in two distinct commercial production sites, one in south of Greece (Galaxidi Marine Farm) and another in southeast of Spain (Cudomar). For this GxE experiment, a population of juveniles was produced by mass spawning of 33 males and 20 females on a single day. These juveniles were stocked in sea cages in both locations when they reached stocking size (~3 g) and grown under commercial conditions. Management conditions during the grow-out period were kept the same between the production sites, while the fish were subject to naturally occurring differences such as water temperature, dissolved oxygen, and salinity. Phenotypes were recorded when the fish reached commercial harvest size (~400 g). Genetic parameters were estimated by using a genomic relationship matrix that was built by using ~30k SNP. All traits studied had higher genetic variation and heritabilities in Cudomar. For instance, the heritability of harvest weight was 0.37 ± 0.05 in Galaxidi and 0.55 ± 0.05 in Cudomar. GxE was estimated as genetic correlations between the same trait measured on different fish in the two environments. Moderate GxE was found for harvest weight (0.45 ± 0.11), growth (0.43 ± 0.11), fillet weight (0.49 ± 0.12), liver weight (0.61 ± 0.11), and viscera weight (0.62 ± 0.10). Weak GxE was found for fillet fat (0.87 ± 0.06), heart weight (0.76 ± 0.11), cardiosomatic index (0.93 ± 0.14), viscerosomatic index (0.90 ± 0.05), and hepatosomatic index (0.79 ± 0.09). In conclusion, moderate GxE estimates for growth traits indicate that with a single breeding program, performance data from both environments should be included, or that two separate breeding programs may be needed for the two environments. The higher genetic variances observed in Cudomar

suggest that this environment is a more suitable test environment for selective breeding.

5.1 Introduction

Animal breeding strives to create populations that will perform well under commercial production circumstances. The performance of genotypes changes in response to varying environmental conditions. Genotype by environment interaction (GxE) occurs when different genotypes respond differently to variation in environmental conditions (Falconer and Mackay, 1996). If GxE leads to reranking of genotypes, the genetically improved animals that perform well in one environment may not perform as expected in another, which implies that the information collected in one environment is of limited value to another environment.

A significant GxE interaction between sites will reduce the effectiveness of a breeding program. The actual production performance of the genetically improved stock will differ from the expected performance based on data from the selection environment. Significant GxE was reported in various production systems, for example, in *Penaeus monodon* culture between indoor recirculating aquaculture systems and outdoor ponds (Sang et al., 2020), in rainbow trout (*Oncorhynchus mykiss*) culture between fresh and brackish water environments (Sae-Lim et al., 2013), and in Nile tilapia (*Oreochromis niloticus*) culture between cage and pond environments (Khaw et al., 2012). Significant GxE was also reported in response to differences in environmental conditions in the same production systems, such as, dissolved oxygen levels (Mengistu et al., 2020a), different nutrition conditions (Romana-Eguia and Doyle, 1992), and salinity levels (Domingos et al., 2021). The issue of GxE attracts more attention when breeding companies start to distribute genetically improved stock internationally (Mulder and Bijma, 2005), in which case the genetically improved animals are expected to perform well in different locations with different production conditions. In aquaculture, one of the species whose genetically improved fingerlings are distributed internationally is gilthead seabream (*Sparus aurata*).

Gilthead seabream is a key fish species for farming in the Mediterranean region. The total production of gilthead seabream increased by 90% between 2009 –

2019 to approximately 260000 tons, making it the second most produced fish species in 2019 in the Mediterranean region (FAO, 2021). In 2019, gilthead seabream was farmed in 23 countries and the leading producers were Turkey, Greece, and Egypt (FAO, 2021). Gilthead seabream is farmed both semi-extensively in ponds and intensively in sea cages or indoor tanks (Elalfy et al., 2021; Lee-Montero et al., 2015). The production of gilthead seabream is spread over large distances from around the Canary Islands in the North Atlantic Ocean (Urbieta and Ginés, 2000) to the shores of the Red Sea in the Arabian Peninsula (FAO, 2017). Although gilthead seabream is farmed in a large geographical area and under various production circumstances, more than 80% of the genetically improved fingerlings originate from a single country, Greece (Janssen et al., 2015), which poses a potential risk for GxE interaction.

If the magnitude of GxE between different production sites is quantified, breeding programs may adjust their selection criteria to reach a balanced performance in many locations rather than a high performance in a single location (Sae-Lim, 2013). However, studies that quantify GxE for commercially important traits in gilthead seabream are limited (Elalfy et al., 2021; Lee-Montero et al., 2015; Navarro et al., 2009a, 2009b). The objective of this study was to quantify GxE for production traits and organ weights of gilthead seabream in two distinct commercial production sites, one in south of Greece and the other in southeast of Spain. The production sites were monitored for changes in water temperature, dissolved oxygen, and salinity. The traits analyzed were harvest weight, growth, fillet weight, fillet percentage, fillet fat percentage, viscera weight, viscerosomatic index, heart weight, cardiosomatic index, liver weight, and hepatosomatic index.

5.2 Materials & Methods

5.2.1 Production sites, production of experimental fish, and grow-out management

Two distinct commercial production sites of gilthead seabream were selected to perform this GxE experiment. One production site is located in south of Greece in

The Gulf of Corinth (Galaxidi Marine Farm S.A., Galaxidi, Greece) (GPS location: 38°21′06.6″N 22°23′18.8″E) and the other production site is located in southeast of Spain at open sea 3.1 km from shore near El Campello, Alicante (Cudomar SL, El Campello, Alicante, Spain) (GPS location: 38°25′12″N 0°20′51″W). For this GxE experiment, a population of juveniles was produced by mass spawning of 33 males and 20 females on a single day at Galaxidi Marine Farm. Juvenile fish were raised as a single group to avoid introduction of common environmental effects. When fish reached an average weight of 3 grams, one batch of 99000 juveniles was stocked in a sea cage at Galaxidi at a density of 0.13 kg/m^3 and another batch of 84605 juveniles was stocked in a sea cage at Cudomar at a density of 0.42 kg/m^3. The fish in both locations were fed with the same commercial diet after reaching an average weight of 100 grams. The feed was provided once a day. The amount of feed given per kilogram of biomass was recorded throughout the grow-out period (Table 5.1).

Table 5.1. Feed given daily (g) per kg of biomass for different size classes.

Average weight of the fish (g)	Feed given (g) per kg of biomass	
	Galaxidi	Cudomar
<100	25.6	29.7
100 - 200	8.9	6.8
200 - 300	10.1	12.1
300 - 400	12.2	10.9

The management conditions during the grow-out period were kept as consistent and similar as possible between the production sites by having meetings between the production managers of the two sites and the researchers before the start of the project. Consensus decisions were made on stocking density based on the availability of sea cages at the time of stocking and on feeding rates based on the effects of water temperature on feeding level throughout the experiment. The environmental conditions of water temperature, dissolved oxygen, and salinity were measured regularly at both locations (Table 5.2). The

dissolved oxygen was measured inside the sea cages in Galaxidi and outside of the sea cages in Cudomar.

Table 5.2. Average environmental and management conditions in Galaxidi and Cudomar.

Variable	Galaxidi	Cudomar
Overall water temperature (°C) [A]	20.5	21.4
Summer water temperature (°C) [A]	25.1	26.1
Autumn water temperature (°C) [A]	20.0	21.7
Winter water temperature (°C) [A]	14.3	15.2
Spring water temperature (°C) [A]	16.5	17.7
Total degree days from stocking until harvest (°C)	3830	4630
Dissolved oxygen (mg/L) [A]	5.58*	8.19*
Salinity (‰)	39.0[B]	37.3[A]

[A] Average daily values, [B] Based on historic steady values

* Dissolved oxygen measurements were made inside of the sea cages in Galaxidi, outside of the sea cages in Cudomar.

The minimum temperature required for gilthead seabream to grow is 12 °C (Hernández et al., 2003). Therefore, the effective daily temperatures were calculated by subtracting 12 °C from observed daily temperatures. The cumulative degree days during the grow-out period were calculated by summing all the effective daily temperatures (Figure 5.1).

5.2.2 Phenotypic data collection

Data on production traits and organ weights were collected from the commercially produced fish that were harvested after a grow-out period of 465 days in Galaxidi and 500 days in Cudomar. Both locations comply with local regulations to produce and harvest seabream in their facilities. In total, 998 fish in Galaxidi and 945 fish in Cudomar were sampled for data collection. The data were collected over 7 days in Galaxidi (daily 100 – 170 fish) and 5 days in Cudomar (daily 150 – 200 fish). A random sample of fish was harvested every morning from the same cage into an oxygenated tank. Before processing, a small batch

of fish received a mortal dose of clove oil (0.03 mL/L). This process was repeated until all the harvested fish were measured for all the traits.

After the fish was killed, body weight was measured with a scale sensitive to 0.5 grams. Fillet fat measurements were taken on whole fish from eight points (four on each side of the fish) by using Distell Fish Fat meter equipment (Distell Inc., West Lothian, Scotland). The fillet fat percentage for each fish was calculated by taking the average of eight measurements on the same fish. The fish were then gutted, and viscera weight was recorded with a scale sensitive to 0.5 grams. Viscera included all the internal organs and abdominal fat. Liver and heart were subsequently separated from the viscera and weighed by a scale sensitive to 0.001 grams. The gutted fish were then filleted, and fillet weight (one side) was recorded. The recorded fillet weight was multiplied by two to calculate the total fillet weight. Measurements were standardized between the two sites, except fillet weight which was skin-off, trimmed in Galaxidi and skin-on, not-trimmed in Cudomar. Measurements taken in a single day were performed by the same person for each trait.

Editing of the data was performed by using R software and the "tidyverse" package collection in R (R Core Team, 2020; Wickham et al., 2019). Fillet percentage was calculated as $(fillet\ weight/body\ weight)\ x\ 100$. The viscerosomatic index was calculated as $(viscera\ weight/body\ weight)\ x\ 100$, the cardiosomatic index was calculated as $(heart\ weight/body\ weight)\ x\ 100$, and the hepatosomatic index was calculated as $(liver\ weight/body\ weight)\ x\ 100$.

Thermal growth coefficient (TGC) is a standardized growth rate that accounts for initial body weight and the sum of daily effective temperatures until harvest. Thermal growth coefficient was calculated as $TGC = \left[\left(W_t^{2/3} - W_0^{2/3}\right)/(T\ x\ t)\right]\ x\ 1000$, where W_t is harvest weight, W_0 is stocking weight, T is the average effective temperature in °C, t is the number of days during grow-out period (Mayer et al., 2012). Stocking weight was on average 2.73 grams in Galaxidi and 3 grams in Cudomar. Because stocking weight was not individually

measured, it was set to 2.73 grams and 3 grams for every fish in Galaxidi and Cudomar, respectively.

5.2.3 DNA extraction and genomic relationship matrix

DNA was isolated from fin clips by IdentiGEN (Dublin, Ireland). The genotyping was performed by using the ~30k "MedFish" SNP array (Peñaloza et al., 2020). After removing 18 duplicate and 24 missing samples, 963 and 939 fish were available for genetic analyses from Galaxidi and Cudomar, respectively.

Genotypic data was filtered by excluding SNPs that had missing call rates exceeding 10%, that were fixed, or had Hardy-Weinberg equilibrium exact test p-value below 1e-10. We computed a genomic relationship matrix (GRM) based on the remaining 28164 SNPs by using calc_grm software (Calus and Vandenplas, 2019). The calculation of the GRM was based on the "vanraden" option (VanRaden, 2008), in which the GRM is computed as $G = \frac{ZZ'}{2\sum_{p_i} p_i\,(1-p_i)}$. To use this option, marker genotypes were coded as "0", "1", or "2". Z is a matrix that contains marker genotypes for all loci, which is corrected for the allele frequency per locus. p_i is the frequency of the less frequent allele and was calculated by using all the fish in both locations. In total, 18 animals were genotyped in duplicate. The GRM was recomputed after removing these animals from the dataset. The inverse of the GRM matrix was obtained directly from calc_grm by using "giv" function. Genotypes of the parents were not available, which prevented the reconstruction of family relationships from parentage analysis.

5.2.4 Estimation of heritabilities and correlations

The magnitude of the GxE was estimated for each trait as the genetic correlation (r_g) of the same trait measured on different animals in the two environments by using a bivariate animal model in ASReml version 4.1 (Gilmour et al., 2015). In the bivariate models, the same traits measured on different fish in the two locations were treated as different traits. Heritabilities of the trait in both environments were estimated by using the same bivariate animal model.

The bivariate animal model is $y = X\beta + Zu + e$, where y is a vector of phenotypes for the same trait in two environments, β is the vector of fixed effect "sampling day", u is the vector of random animal additive genetic effects $\sim \left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, G \begin{bmatrix} \sigma^2_{a,T_G} & r_{a,T_{G,C}} \sigma_{a,T_G} \sigma_{a,T_C} \\ r_{a,T_{G,C}} \sigma_{a,T_G} \sigma_{a,T_C} & \sigma^2_{a,T_C} \end{bmatrix} \right)$, where G is the genomic relationship matrix and σ^2_{a,T_G} is the additive genetic variance of trait measured in Galaxidi, σ^2_{a,T_C} is the additive genetic variance of trait measured in Cudomar, $r_{a,T_{G,C}}$ is the additive genetic correlation between the same trait measured in Galaxidi and Cudomar and e is the vector of random residual effects $\sim \left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, I \begin{bmatrix} \sigma^2_{e,T_G} & 0 \\ 0 & \sigma^2_{e,T_C} \end{bmatrix} \right)$, where I is an identity matrix, σ^2_{e,T_G} is the residual variance of trait measured in Galaxidi, and σ^2_{e,T_C} is the residual variance of trait measured in Cudomar. The residual covariance between the two environments was set to zero because no individual was measured both in Galaxidi and Cudomar. X and Z are design matrices, that relate observations to the fixed effect and additive genetic effect of animals, respectively. The fixed effect "sampling day" was a categorical variable with one category for each measurement day (7 days in Galaxidi and 5 days in Cudomar). Gilthead seabream is a protandrous fish (Loukovitis et al., 2011) and all individuals were males during harvest. Therefore, "sex" was not included as a fixed effect in this study. The expectation of $cov(u, e)$ is zero. Heritability (h^2) of each trait was calculated as $h^2 = \frac{\sigma^2_a}{\sigma^2_a + \sigma^2_e}$. After fitting the bivariate animal models for each trait by using all 963 and 939 phenotypic records in Galaxidi and Cudomar, residuals that were more than 3.5 standard deviations in magnitude were identified and the corresponding phenotypic records were removed. Table 3 shows the number of fish used in the analysis of each trait.

Genetic and phenotypic correlations among the traits within the same environment were also estimated by using bivariate animal models. In this case, u is the vector of random animal additive genetic effects

$\sim\left(\begin{bmatrix}0\\0\end{bmatrix}, G\begin{bmatrix}\sigma^2_{a,T_1} & r_{a,T_{1,2}}\sigma_{a,T_1}\sigma_{a,T_2}\\ r_{a,T_{1,2}}\sigma_{a,T_1}\sigma_{a,T_2} & \sigma^2_{a,T_2}\end{bmatrix}\right)$, where σ^2_{a,T_1} is the additive genetic variance of trait 1, σ^2_{a,T_2} is the additive genetic variance of trait 2, $r_{a,T_{1,2}}$ is the additive genetic correlation between trait 1 and 2. Also, $\boldsymbol{e}$ is the vector of random residual effects $\sim\left(\begin{bmatrix}0\\0\end{bmatrix}, I\begin{bmatrix}\sigma^2_{e,T_1} & r_{e,T_{1,2}}\sigma_{e,T_1}\sigma_{e,T_2}\\ r_{e,T_{1,2}}\sigma_{e,T_1}\sigma_{e,T_2} & \sigma^2_{e,T_2}\end{bmatrix}\right)$, where I is an identity matrix, σ^2_{e,T_1} is the residual variance of trait 1, and σ^2_{e,T_2} is the residual variance of trait 2, and, $r_{e,T_{1,2}}$ is the residual correlation between trait 1 and trait 2.

5.3 Results

5.3.1 Descriptive statistics

Descriptive statistics of body weight at harvest, thermal growth coefficient, fillet weight, fillet percentage, fillet fat percentage, viscera weight, viscerosomatic index, liver weight, hepatosomatic index, heart weight, and cardiosomatic index were calculated (Table 5.3).

Table 5.3. The number of fish (N) used for analyses and, means and coefficients of variation (CV) of all traits in the two production environments.

Trait	Galaxidi			Cudomar		
	N	Mean*	CV	N	Mean*	CV
Harvest weight (g)	957	372.1	17.0	938	412.2	17.0
TGC ($g^{2/3}$ x $°C^{-1}$ x 1000)	955	12.8	11.9	936	11.4	11.9
Fillet weight (g)	960	116.6	19.6	933	179.8	19.8
Fillet percentage (%)	946	31.4	8.1	929	43.6	7.1
Fillet fat (%)	950	12.7	18.1	932	12.8	21.6
Viscera weight (g)	963	27.2	27.0	935	29.3	23.1
Viscerosomatic index (%)	961	7.3	16.8	934	7.1	13.9
Liver weight (g)	947	3.9	29.1	930	6.8	31.4
Hepatosomatic index (%)	945	1.0	21.2	933	1.6	23.4
Heart weight (g)	963	0.38	25.6	933	0.48	21.4
Cardiosomatic index (%)	954	0.10	21.6	927	0.12	13.4

*All trait means were significantly different between sites ($p < .001$) except for fillet fat (%).

The harvested fish in Cudomar were on average 40.1 grams heavier than in Galaxidi; however, the coefficients of variation were virtually the same. TGC was higher in Galaxidi. The average fillet weight of harvested fish in Cudomar was on average 63.2 grams higher than in Galaxidi; however, this is mainly because the fillet was skin-on and not trimmed in Cudomar, while it was skin-off and trimmed in Galaxidi. The coefficients of variation of fillet weight were very similar. The viscera weights of harvested fish in the two environments were very similar despite the difference of 40.1 grams in harvest weights. The liver of harvested fish in Cudomar was on average more than 70% heavier than in Galaxidi, which was also reflected in the large difference in hepatosomatic indices. Liver and heart weights, and corresponding indices had high coefficients of variation in both environments. The mean fillet fat percentages of harvested fish in the two environments were almost identical; however, the coefficient of variation was somewhat higher in Cudomar.

5.3.2 Genetic parameters

The only fixed effect in our dataset, "sampling day", was significant in the analysis of all traits ($p < .05$) except for harvest and viscera weight in Galaxidi. The estimates of genetic variance, environmental variance, and heritability were obtained from bivariate models for each trait (Table 5.4).

The heritability estimates were higher for all traits in Cudomar than in Galaxidi, although the differences were significant only for fillet weight and hepatosomatic index ($p < .05$). The heritability estimates of harvest weight were moderate to high. The heritability estimates of TGC were almost identical to the heritability estimates of harvest weight. Heritability estimates of fillet percentage were low in both environments. In comparison, the heritabilities of viscerosomatic, hepatosomatic, and cardiosomatic indices were higher than fillet percentage in both environments, with higher values in Cudomar. The heritability estimates for liver and heart weight were moderate to high in both environments. The heritability for cardiosomatic index was lower than the one for heart weight, but the difference was not as big as between fillet weight and fillet percentage.

Table 5.4. Genetic variance (V_A), environmental variance (V_E), and heritability of the traits in the two production environments.

Trait	Galaxidi			Cudomar		
	V_A	V_E	h^2 (se)	V_A	V_E	h^2 (se)
Harvest weight	1474	2553	0.37 (0.05)	2863	2361	0.55 (0.05)
TGC	0.87	1.49	0.37 (0.05)	1.09	0.87	0.56 (0.05)
Fillet weight	34.60	82.7	0.30 (0.05)	156.32	156.10	0.50 (0.05)
Fillet percentage	0.09	1.21	0.07 (0.04)	0.21	1.49	0.13 (0.04)
Fillet fat (%)	2.24	2.64	0.46 (0.05)	3.54	2.94	0.55 (0.05)
Viscera weight	21.84	33.12	0.40 (0.05)	23.16	26.24	0.47 (0.05)
VSI	0.66	0.85	0.44 (0.05)	0.50	0.48	0.51 (0.05)
Liver weight	0.43	0.84	0.34 (0.05)	2.05	2.48	0.45 (0.05)
HSI	0.013	0.032	0.29 (0.05)	0.066	0.070	0.48 (0.05)
Heart weight	0.002	0.007	0.23 (0.05)	0.004	0.006	0.39 (0.05)
CSI	0.69	3.56	0.16 (0.04)	0.48	2.02	0.19 (0.05)

VSI = Viscerosomatic index, HSI = Hepatosomatix index, CSI = Cardiosomatic index

5.3.3 Genetic and phenotypic correlations

The genetic and phenotypic correlations between traits were estimated with fish grown in Galaxidi (Table 5.5) and in Cudomar (Table 5.6). The genetic correlation between harvest weight and TGC was close to unity in both environments, indicating these two traits are genetically same. The genetic correlation between harvest weight and fillet weight was also very close to unity, which means selecting for increased harvest weight will result in correlated favorable response in fillet weight. However, the genetic correlation between harvest weight and fillet percentage was negative in Galaxidi (although the correlation was weak and not significantly different from zero) and positive in Cudomar (with a large standard error). This indicates that selecting for increased harvest weight in Galaxidi will not result in correlated favorable response in fillet percentage; however, some correlated response is expected in Cudomar. Harvest weight was strongly and positively correlated with viscera and organ weights, which means that selecting for increased harvest weight will result in correlated response for increased organ weights; however, the genetic correlations between harvest weight and

Table 5.5. Genetic (below diagonal) and phenotypic (above diagonal) correlations among the traits measured at harvest in Galaxidi. HW (harvest weight), TGC (thermal growth coefficient), FW (fillet weight), FP (fillet percentage), FF% (fillet fat percentage), VW (viscera weight), VSI (viscerosomatic index), LW (liver weight), HSI (hepatosomatic index), HeW (heart weight), and CSI (cardiosomatic index).

	HW	TGC	FW	FP	FF%	VW	VSI	LW	HSI	HeW	CSI
HW		>0.99 (<0.01)	0.90 (0.01)	0.05 (0.04)	0.32 (0.04)	0.79 (0.02)	0.28 (0.04)	0.69 (0.02)	0.18 (0.04)	0.56 (0.03)	-0.16 (0.04)
TGC	>0.99 (<0.01)		0.91 (0.01)	0.09 (0.04)	0.38 (0.04)	0.80 (0.02)	0.31 (0.04)	0.70 (0.02)	0.21 (0.04)	0.57 (0.03)	-0.18 (0.04)
FW	0.98 (0.01)	0.98 (0.01)		0.46 (0.03)	0.41 (0.03)	0.73 (0.02)	0.28 (0.04)	0.67 (0.02)	0.23 (0.04)	0.52 (0.03)	-0.16 (0.04)
FP	-0.15 (0.22)	-0.10 (0.25)	0.11 (0.25)		0.26 (0.03)	0.06 (0.04)	0.06 (0.04)	0.11 (0.04)	0.13 (0.04)	0.05 (0.03)	0.13 (0.04)
FF%	0.16 (0.13)	0.17 (0.13)	0.24 (0.13)	0.38 (0.19)		0.34 (0.04)	0.23 (0.04)	0.48 (0.03)	0.43 (0.03)	0.11 (0.04)	-0.19 (0.04)
VW	0.75 (0.05)	0.77 (0.05)	0.75 (0.06)	-0.04 (0.24)	0.26 (0.12)		0.80 (0.01)	0.70 (0.02)	0.35 (0.03)	0.43 (0.03)	-0.16 (0.04)
VSI	0.24 (0.11)	0.27 (0.11)	0.26 (0.12)	0.01 (0.22)	0.29 (0.11)	0.82 (0.04)		0.42 (0.03)	0.37 (0.04)	0.14 (0.04)	0.01 (0.04)
LW	0.73 (0.07)	0.76 (0.06)	0.77 (0.07)	-0.02 (0.25)	0.50 (0.10)	0.67 (0.07)	0.35 (0.11)		0.83 (0.01)	0.36 (0.03)	-0.17 (0.04)
HSI	0.25 (0.13)	0.27 (0.13)	0.29 (0.14)	0.05 (0.26)	0.64 (0.09)	0.36 (0.12)	0.28 (0.12)	0.83 (0.04)		0.06 (0.04)	-0.11 (0.04)
HeW	0.81 (0.07)	0.81 (0.07)	0.75 (0.09)	-0.28 (0.27)	-0.08 (0.15)	0.46 (0.12)	-0.04 (0.14)	0.48 (0.13)	-0.02 (0.17)		0.69 (0.02)
CSI	-0.11 (0.18)	-0.11 (0.19)	-0.17 (0.19)	0.05 (0.26)	-0.35 (0.16)	-0.34 (0.16)	-0.16 (0.16)	-0.35 (0.18)	-0.41 (0.18)	0.50 (0.14)	

Table 5.6. Genetic (below diagonal) and phenotypic (above diagonal) correlations among the traits measured at harvest in Cudomar. HW (harvest weight), TGC (thermal growth coefficient), FW (fillet weight), FP (fillet percentage), FF% (fillet fat percentage), VW (viscera weight), VSI (viscerosomatic index), LW (liver weight), HSI (hepatosomatic index), HeW (heart weight), and CSI (cardiosomatic index).

	HW	TGC	FW	FP	FF%	VW	VSI	LW	HSI	HeW	CSI
HW		>0.99 (<0.01)	0.95 (0.01)	0.20 (0.04)	0.28 (0.04)	0.81 (0.02)	0.18 (0.04)	0.71 (0.02)	0.25 (0.04)	0.77 (0.02)	-0.06 (0.04)
TGC	>0.99 (<0.01)		0.95 (0.01)	0.21 (0.04)	0.30 (0.05)	0.81 (0.02)	0.19 (0.04)	0.71 (0.02)	0.26 (0.04)	0.77 (0.02)	-0.06 (0.04)
FW	0.99 (0.01)	0.99 (0.01)		0.48 (0.03)	0.33 (0.04)	0.76 (0.02)	0.15 (0.04)	0.69 (0.02)	0.28 (0.04)	0.74 (0.02)	-0.04 (0.04)
FP	0.39 (0.15)	0.39 (0.15)	0.50 (0.13)		0.28 (0.03)	0.14 (0.04)	-0.01 (0.04)	0.23 (0.17)	0.19 (0.04)	0.17 (0.04)	0.02 (0.04)
FF	0.14 (0.10)	0.15 (0.10)	0.20 (0.10)	0.62 (0.13)		0.32 (0.04)	0.21 (0.04)	0.46 (0.03)	0.45 (0.03)	0.15 (0.04)	-0.09 (0.04)
VW	0.80 (0.04)	0.80 (0.04)	0.72 (0.05)	0.30 (0.18)	0.23 (0.10)		0.70 (0.02)	0.76 (0.02)	0.46 (0.04)	0.65 (0.02)	-0.03 (0.04)
VSI	0.06 (0.11)	0.06 (0.11)	-0.08 (0.11)	-0.43 (0.15)	0.25 (0.10)	0.61 (0.07)		0.43 (0.04)	0.85 (0.01)	0.16 (0.04)	0.01 (0.04)
LW	0.68 (0.06)	0.69 (0.06)	0.64 (0.07)	0.29 (0.17)	0.45 (0.09)	0.81 (0.04)	0.45 (0.09)		0.85 (0.01)	0.53 (0.03)	-0.07 (0.04)
HSI	0.15 (0.10)	0.16 (0.10)	0.12 (0.11)	0.22 (0.17)	0.54 (0.08)	0.45 (0.09)	0.54 (0.08)	0.81 (0.04)		0.19 (0.04)	-0.05 (0.04)
HeW	0.92 (0.03)	0.92 (0.03)	0.95 (0.03)	0.54 (0.16)	-0.06 (0.12)	0.70 (0.06)	-0.01 (0.12)	0.55 (0.09)	0.06 (0.12)		0.58 (0.03)
CSI	-0.02 (0.15)	-0.01 (0.15)	0.08 (0.16)	0.49 (0.22)	-0.26 (0.14)	-0.08 (0.16)	-0.16 (0.16)	-0.15 (0.16)	-0.19 (0.15)	0.39 (0.13)	

organ indices were not significantly different than zero, indicating that selecting for increased harvest weight will not result in correlated response for the increased proportion of organs to body weight.

Fillet fat percentage was positively correlated to viscera weight and viscerosomatic index, indicating that selecting for heavier viscera will result in correlated response for increased fillet fat. Fillet fat percentage was also positively correlated to liver weight and hepatosomatic index, which means that selecting for increased liver weight will result in correlated response for increased fillet fat.

5.3.4 Genotype by environment interaction between the production environments

Genotype by environment interactions were estimated as genetic correlations between the same traits in two production environments (Table 5.7).

Table 5.7. Genetic correlations between the production environments of Galaxidi and Cudomar for all traits.

Trait	Genetic correlation (se)
Harvest weight	0.45 (0.11)
TGC	0.43 (0.11)
Fillet weight	0.49 (0.12)
Fillet percentage	0.51 (0.30)
Fillet fat (%)	0.87 (0.06)
Viscera weight	0.62 (0.10)
Viscerosomatic index	0.90 (0.05)
Liver weight	0.61 (0.11)
Hepatosomatic index	0.79 (0.09)
Heart weight	0.76 (0.11)
Cardiosomatic index	0.93 (0.14)

TGC had the lowest genetic correlation between the two production environments, and therefore the strongest genotype by environment interaction.

Harvest weight and TGC had very similar levels of genotype by environment interaction. The GxE interaction for fillet weight, liver weight, and viscera weight was moderate. On the other hand, the GxE interaction for fillet fat percentage, heart weight, cardiosomatic index, and viscerasomatic index was weak.

5.4 Discussion

The objective of this study was to quantify genotype by environment interaction for production traits and organ weights of gilthead seabream in two distinct production sites in the Mediterranean, one site in The Gulf of Corinth, Greece (Galaxidi Marine Farm) and the other site in open sea near El Campello, Spain (Cudomar). Estimates for heritabilities were higher in Cudomar. TGC, harvest weight and fillet percentage showed the strongest genotype by environment interaction, which are traits of economic importance and commonly included in fish breeding programs (Chavanne et al., 2016; Janssen et al., 2017). Fillet fat percentage, on the other hand, had weak genotype by environment interaction.

5.4.1 Experimental design

The two commercial production locations were specifically chosen to analyze the effect of distinct temperature profiles on genotype by environment interaction with respect to production traits and organ weights. The historical sea surface temperature data indicate that the sea surface water in south of Greece is warmer than in southeast of Spain, which makes Greece the more favorable environment for growth (Besson et al., 2016; Llorente and Luna, 2013). Water temperatures recorded during this experiment however, revealed a different pattern (Figure 5.1), which was mainly due to the specific location of Galaxidi, which is in The Gulf of Corinth. In this experiment, Cudomar was characterized by consistently higher daily average water temperatures during the first summer, which led to increasing differences in cumulative degree days in favor of Cudomar until the beginning of winter. Winter only fortified the divergence of

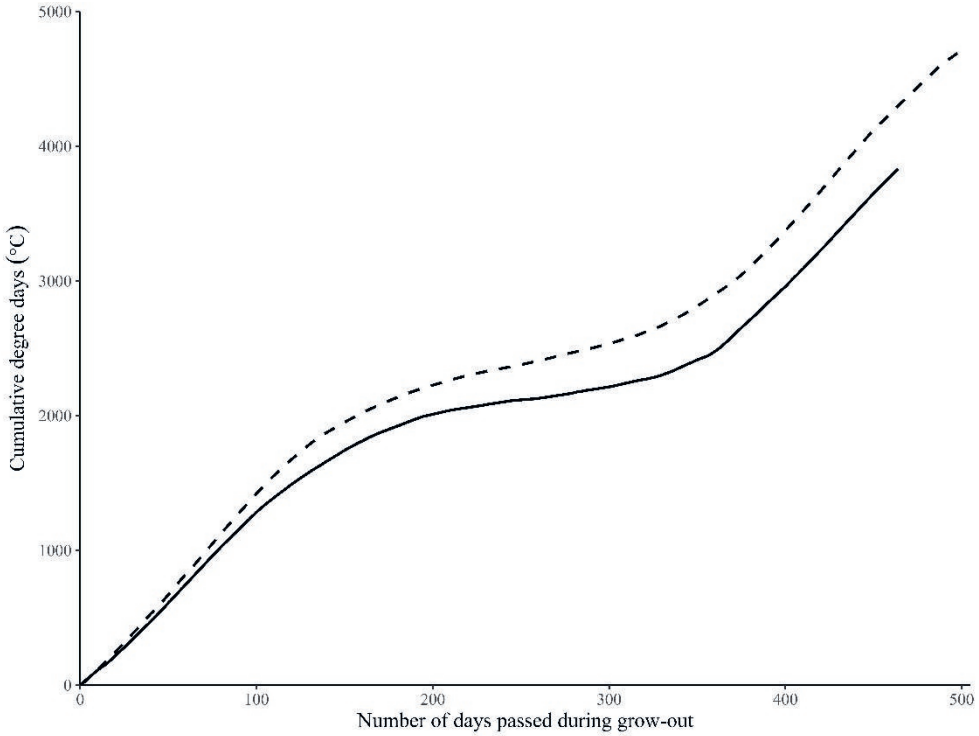

Figure 5.1. Cumulative degree days (°C) in Galaxidi (——) and Cudomar (- - - -) from stocking to harvest.

cumulative degree days until the beginning of second summer, which did not cause further divergence in cumulative degree days. The 35 days gap in harvesting dates was also an influential factor on the final difference on cumulative degree days. Water temperature has a direct effect on the performance of fish. Fish that are subjected to higher temperatures grow faster (Green and Fisher, 2004); however, temperatures above certain limits inhibit the growth (Azaza et al., 2008). The water temperature in the Mediterranean has been increasing consistently over the last decades (Pastor et al., 2018). The temperature has increased at different levels in different parts of the Mediterranean. Therefore, increasing water temperature is likely to become a major factor contributing to genotype by environment interaction in fish farming

in the future as temperature has a direct effect on physiological performance. In addition to water temperature, the two locations differed in the levels of dissolved oxygen and salinity (Table 5.1), which may also have contributed to genotype by environment interactions. Lower salinity was concluded to result in improved growth in various marine fish species including gilthead seabream (McKay and Gjerde, 1985; Morgan and Iwama, 1991; Tandler et al., 1995; Ytrestøyl et al., 2020). The underlying reason for this may be that higher salinity demands higher metabolic costs to maintain the osmotic pressure of the body fluids within acceptable ranges, therefore any saved energy due to lower salinity can be spent for growth (Morgan and Iwama, 1991; Tandler et al., 1995). Feed efficiency was detected to improve in lower salinity, which reflects that more of the feed is used for growth rather than for other metabolic processes (Imsland et al., 2001; Ytrestøyl et al., 2020). Higher dissolved oxygen concentrations were reported to improve growth and feed efficiency (Duan et al., 2011; Mengistu et al., 2020a, 2020b). Under hypoxia, the blood flow through the intestines is restricted (Axelsson and Fritsche, 1991), which can negatively affect nutrient absorption and feed utilization. Taking these observations together, Cudomar appeared to be the more favorable environment for fish growth, which could explain why the estimates of genetic variation (Table 5.4) were higher in Cudomar compared to Galaxidi.

In this study, genotype by environment interaction was quantified as the genetic correlation between the same trait measured in the two production environments. The experimental fish were grown in commercial sea cages to commercial harvest size, and they were subject to natural environmental conditions. In this study, no systematic manipulation of environmental conditions was applied, and the same commercial feed was given in both sites from an average weight of 100 grams until the time of harvest.

The fish that were stocked in sea cages in Galaxidi and Cudomar resulted from a single mass spawning event. In mass spawning, it is not possible to control the

mating and the contribution of the parents may be skewed (Loughnan et al., 2013). While mass spawning leads to unequal contribution of parents, which is a disadvantage for genetic parameter estimation, it also avoids introduction of common environmental effects that could obscure genetic differences between families. Unequal family sizes may create a slight bias on the estimations of genotype by environment interaction when the analyses are performed using the pedigree relationships (Sae-Lim et al., 2010); however, the genetic analyses were performed using genomic relationships in this study. The use of genomic relationships increases the accuracy of parameter estimates as compared to the use pedigree relationships (Veerkamp et al., 2011).

5.4.2 Heritability of the traits

In this study, the genetic variances of all the traits were higher in Cudomar, which also resulted in higher heritability estimates. One of the possible consequences of genotype by environment interaction is heterogenous genetic variances across environments (Calus, 2006). The heterogenous genetic variances between the two environments (Table 5.4) lead to scaling effects. It appears that the environmental conditions (both measured and unmeasured) in Cudomar were more favorable than in Galaxidi, which led to higher genetic variances. This may imply that when breeding for these two environments, keeping the selection candidates at Cudomar may increase the selection response due to the higher genetic variances observed in Cudomar. Higher heritabilities would also lead to higher accuracies of EBVs.

Harvest weight was moderately heritable in Galaxidi and highly heritable in Cudomar. The heritability of harvest weight in Galaxidi (0.37 ± 0.05) is very similar to the value reported for gilthead seabream (0.34 ± 0.06) by Navarro et al., (2009), although their estimation comes from a population that consisted of a combination of tank-farmed and cage-farmed fish. The value reported by Fernandes et al., (2017) (0.41 ± 0.03) for harvest weight of tank-farmed gilthead seabream is in between the values estimated for Galaxidi (0.37 ± 0.05) and

Cudomar (0.55 ± 0.05). The genetic correlation between harvest weight and TGC was almost one, which indicates that these traits are controlled by the same set of genes. The heritability estimates of harvest weight and TGC were almost equal. These may be related to how the TGC was calculated. TGC depends on harvest weight, stocking weight, water temperature, and duration of the grow-out period. In this study, individual stocking weights were not available, and an average was used for all fish in the same location. The stocking weights, however, did not affect TGC to a high degree because they were very small, less than one percent of the harvest weights. Therefore, the variation in TGC was dominated by the variation in harvest weight.

Fillet percentage was lowly heritable in both production environments (0.07 ± 0.04 in Galaxidi and 0.13 ± 0.04 in Cudomar), similar to the value reported by Navarro et al., (2009a) (0.12 ± 0.03) for a population that consisted of a combination of tank-farmed and cage-farmed fish and slightly lower than the estimate reported by Vandeputte et al., (2020) (0.22 ± 0.05) for cage-farmed gilthead seabream. The heritability estimates of fillet percentage are typically found to be lower than the heritability of fillet weight (Navarro et al., 2009a; Thodesen et al., 2012; Vandeputte et al., 2017), which is also the case in this study. The differences in heritability estimates for fillet percentage may partly be caused by different filleting methods (skin-off, trimmed in Galaxidi and skin-on, not trimmed in Cudomar) used in the two production environments (Kocour et al., 2007; Thodesen et al., 2012). The genetic coefficient of variations for fillet percentage were very small, which indicates that the prospect of genetic improvement of this trait by direct selection is not good. Fillet percentage may be improved to some extent by selecting on harvest weight based on the positive genetic correlation between harvest weight and fillet percentage in Cudomar.

Fillet fat percentage was highly heritable in both production environments (0.46 ± 0.05 in Galaxidi and 0.55 ± 0.05 in Cudomar). The heritability of fillet fat estimated in this study is larger than the value reported by Elalfy et al. (2021)

(0.27 ± 0.08), who also measured fillet fat of gilthead seabream with Distell equipment. The heritability of fillet fat estimated in this study is also larger than the values reported for striped catfish (0.04 ± 0.03) (Sang et al., 2012), for European whitefish (0.26 ± 0.09) (Kause et al., 2011) and for Atlantic salmon (0.28 ± 0.05) (Powell et al., 2008); however, lower than the value reported for common carp (0.68 ± 0.10) (Prchal et al., 2018). High genetic coefficients of variation were found for fillet fat, which indicate that this trait would respond well to genetic selection. Fillet fat content may be regarded as a quality trait of the farmed fish from the perspective of consumers (Sang et al., 2009). Consumers in certain markets may prefer to purchase fish with a higher fillet fat content. Even though gilthead seabream farmers are not paid on the basis of quality characteristics of the meat, farmers may wish to advertise their product as having higher quality due to increased fillet fat to allure a certain market and distinguish themselves in the market. On the other hand, fillet fat is positively (unfavorably) correlated to viscera weight. Viscera is a storage for fat in the body. Fat content of the body is expected to be unfavorably correlated to feed conversion ratio and therefore, selecting for increased fillet fat may create fish that are less efficient converters of feed (Kause et al., 2016; Quinton et al., 2007). Therefore, rather than extreme values, an optimum is required for fillet fat percentage.

Heart and liver weights, and indices were moderately heritable in both production environments, although the estimations were higher in Cudomar. In both environments, the genetic correlation between heart and harvest weights was high but not unity. Therefore, selection for growth could result in fish with slightly lower cardiosomatic indices. A lower cardiosomatic index may be associated with reduced cardiac capacity and consequently, reduced robustness of the fish (Vassgård, 2017). Growth rate of the fish has been reported to affect the prevalence of cardiac diseases (Farrell, 2002), which may be caused by the increased pressure on the heart to supply more blood to the fast-growing body. In broiler chickens, cardiac diseases, specifically ascites, was observed to have

higher prevalence in faster-growing animals (Closter et al., 2009). The increase of ascites in broilers is related to genetic selection for higher performance (Julian, 1993). Therefore, heart weight is an important trait to monitor for unwanted correlated responses that could affect health and robustness. Heart weight is a difficult-to-measure trait because measuring it requires sacrificing the fish. However, genomic selection makes it easier to implement such traits in a breeding program because after phenotyping and genotyping a reference population, the selection candidates do not have to be phenotyped and can still receive accurate breeding values. Hepatosomatic index has been reported to increase in response to differences in lighting regime (Døskeland et al., 2016) and use of additives in feed in Atlantic salmon (*Salmo salar*) (Larsson et al., 2014), exposure to insecticides in Nile tilapia (*Oreochromis niloticus*) (Thomaz et al., 2009), and exposure to toxins in redbreast sunfish (*Lepomis auritus*) (Everaarts et al., 1993). Enlargement of the liver may be positively correlated by its capacity to transform xenobiotics and cleanse the body (Porter and Janz, 2003). Therefore, hepatosomatic index can be regarded as a general health indicator of fish (Thomaz et al., 2009). Suboptimal environmental conditions during grow-out may cause the hepatosomatic indices to increase because the growth of the body is more susceptible to environmental conditions than the growth of the organs (Døskeland et al., 2016). However, the effect of individual environmental factors, such as water temperature and salinity, on hepatosomatic index is unknown.

5.4.3 Genotype by environment interaction and implications for breeding programs

Genotype by environment interaction may manifest itself as scaling or reranking of genotypes. In case of reranking of genotypes, good performing genotypes in one location may not perform as expected in another location. It was suggested that the reranking of genotypes due to genotype by environment interaction is practically important only if the genetic correlation of the same trait in different

environments is below 0.8 (Robertson, 1959). In this experiment, environmental effects that can be controlled, such as feeding and stocking density, were kept as consistent as possible between the two production sites and fish were subject to natural water conditions that cannot be controlled, such as water temperature and salinity. Therefore, any genotype by environment interaction present is most likely due to different water conditions rather than different husbandry practices.

The genetic correlation of harvest weight was 0.45 ± 0.11 and the genetic correlation of TGC was 0.43 ± 0.11 between Galaxidi and Cudomar, which indicates a moderate genotype by environment interaction for these traits. Navarro et al., (2009a) estimated a weaker genotype by environment interaction for harvest weight (0.70 ± 0.10) in gilthead seabream although the experimental fish in that study were reared in different production systems (tanks and sea cages). In tanks, environmental conditions can be controlled, which may then lead to low genotype by environment interaction. Lee-Montero et al. (2015) estimated weak to strong genotype by environment interaction for specific growth rate (0.05 ± 0.18 – 0.99 ± 0.12) in gilthead seabream among different regions of Spain and varying grow-out conditions including intensive rearing in tanks and sea cages, and semi extensive rearing in ponds. In their study, grow-out conditions as well as geographical distance strongly affected the magnitude of genotype by environment interaction. Growth is the primary trait of interest in aquaculture and highly relevant for breeding programs (Chavanne et al., 2016). Growth and harvest weight had very similar genotype by environment interaction in this study. Moderate genotype by environment interaction detected for growth indicates that an optimization procedure may be applied if the interest is in reaching a balanced performance in the two environments. An obvious optimization procedure is to standardize the environmental conditions between the two environments (Sae-Lim et al., 2013); however, husbandry practices in this experiment were largely standardized and sea cages were subject to naturally fluctuating water conditions that cannot be controlled. Other optimization

options are combining the performance of fish in both environments in an index and selecting on this index in a single breeding program or running two separate breeding programs for the two environments (James, 1961). If genotype by environment interaction is present, using performance data from the second production environment in an index increases the genetic gain in that environment significantly (Chu et al., 2018). To decide whether to select on an index or to run separate breeding programs, the "break-even correlation" should be considered. The break-even correlation is the genetic correlation between the same trait in different environments at which the genetic gain of the two approaches is equal when the costs of two smaller breeding programs are the same as the cost of one large breeding program (Mulder et al., 2006). Mulder et al., (2006) suggests applying separate breeding programs if the genetic correlation is below 0.5 – 0.7.

The interest of breeding companies that distribute genetic material widely would be to produce fish that performs well in many conditions and environments. Therefore, depending on the genotype by environment interaction, data of more than two locations may need to be used in a breeding program to reach a balanced performance across those locations. Several methods are available to group different types of environments in a breeding program (Chenu, 2015). The number and type of locations depend on the specific interests of the breeding companies.

Genotype by environment interaction for fillet weight was moderate. The GxE interaction for fillet weight was slightly weaker than for harvest weight, although not significantly different. This small difference is in line with the result of Navarro et al., (2009a) who also estimated weaker genotype by environment interaction for fillet weight (0.94 ± 0.19) than harvest weight (0.70 ± 0.10) in two production sites of gilthead seabream. However, their estimate of 0.94 ± 0.19 was much higher than our estimate (0.49 ± 0.12). The magnitude of genotype by

environment interaction can depend highly on the specific differences between the environments.

The genetic correlation of fillet fat percentage in Galaxidi and Cudomar was 0.87 ± 0.06, which indicates a very weak genotype by environment interaction. This may mean that accumulation of fat in the muscle is not strongly affected by environmental changes. The weak genotype by environment interaction is in accordance with the result of Elalfy et al. (2021), who estimated no genotype by environment interaction for fillet fat (1.00 ± 0.00) between intensive and semi-extensive grow-out conditions. Navarro et al. (2009b) estimated a much stronger genotype by environment interaction for fillet fat (0.15 ± 0.94) measured with chemical methods in gilthead seabream; however, their estimation was with a very high standard error and probably not significantly different from our estimate. A genetic correlation of 0.86 ± 0.06 for fillet fat percentage indicates that reranking of genotypes is minimal to absent for this trait.

5.5 Conclusion

Local water conditions may differ considerably between production sites, which indicates that local niches play an important role in the production of gilthead seabream. In this study, moderate genotype by environment interaction was detected for harvest weight, TGC, fillet weight and liver weight, and weak genotype by environment interaction was detected for fillet fat percentage and heart weight. This shows that the relative genetic merit of gilthead seabream in different farms can vary considerably and that breeding programs need to consider using data from multiple locations if the produced fish must perform in different locations. With the moderate genetic correlations observed, separate breeding programs may be needed.

Acknowledgement

Authors would like to thank Galaxidi Marine Farm and Cudomar for their participation in the study and their staff for providing necessary help and facilities for raising the population and the experimental data collection.

Funding

This project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No 727315.

Chapter 6:

General Discussion

6.1 Introduction

Genetic selection has successfully been contributing to improvement of farmed animal populations in terms of performances and welfare. Breeding programs are systematic approaches to perform genetic selection that aims to identify genetically superior animals to serve as the parents of the next generation. By reproducing genetically superior parents, genetic improvement accumulates in each generation.

In this thesis, I make a distinction between central-nucleus and integrated breeding programs due to two aspects. First aspect is where and how the selection candidates are kept. In central-nucleus approach, selection candidates are typically maintained in bio-secure indoor facilities with optimum environmental conditions whereas in integrated approach, selection candidates are the production animals of fish farms that are kept under specific conditions of the farm's production location. The second aspect is dissemination of genetic improvement. In central-nucleus approach, the genetic improvement is disseminated to fish farms from the "central" location of nucleus and there is typically a time lag for genetic improvement to reach fish farms. In integrated approach, genetic improvement can be utilized in commercial production without any time lag because fish farms use their own hatcheries to produce their own commercial stock from the latest generation of broodstock.

There are many advanced central-nucleus breeding programs for farmed terrestrial animal species. Some of these programs date back to ~100 years ago including a pedigree-based breeding program for broiler chickens (Lee, 2023), for laying hens (Tixier-Boichard et al., 2012), for dairy cattle (Miglior et al., 2017), and for beef cattle (Sasaki et al., 2006). The number of breeding programs is not limited to one for each species as there can be several focusing on fulfilling different needs of farmers, such as for highly efficient or high yielding broiler lines (Cobb-Vantress Inc., 2023). The implementation of central-nucleus breeding programs for improvement of fish species is more recent than terrestrial animals.

The first family-based central-nucleus breeding program started in 1975 for Atlantic salmon (Gjedrem, 2010), which was followed by another program for Nile tilapia that started in 1988 (Bentsen et al., 2017). These are now advanced implementations of central-nucleus approach with genomic selection implemented (Joshi et al., 2021; Robledo et al., 2018). There are several central-nucleus breeding programs for other fish species although their implementation is not as advanced as the two initial ones (Boudry et al., 2021).

For some fish species, both central-nucleus and integrated breeding programs are available; an example is gilthead seabream. The genetic improvement of seabream started in early 2000's (Janssen et al., 2015), and nine breeding programs were reported to be active in 2016 (Chavanne et al., 2016). Naturally, the implementation practices and the degree of genetic improvement are different in those programs. Only a single breeding program has been reported to be the main business activity for its executing company; others have been reported to be "integrated" with commercial fish production, meaning that the companies produce genetically improved juveniles for their own commercial use and the harvested fish generates the main source of income rather than selling genetically improved eggs/juveniles (Janssen et al., 2015). "Integrated" with commercial production does not necessarily mean that the activity is an integrated breeding program. Some seabream farming companies run their own central-nucleus breeding programs to supply genetically improved juveniles for their own commercial production; an example is Nireus SA in Greece. In this case, the selection candidates are a distinct population that are kept at a "central" indoor location inside the fish farm. Some other seabream farming companies run integrated breeding programs to supply juveniles for their own commercial production; an example is Galaxidi Marine Farm SA in Greece, for which I reported genetic parameters for several production traits and organ weights in Chapter 5. Integrated approach might be preferable to central-nucleus approach in certain cases.

In this thesis, I aimed to contribute to the improvement of integrated breeding programs for fish species with innovative research. This thesis includes four research chapters, each of which addresses a specific challenge faced by integrated breeding programs. Chapter 2 addresses the challenge to obtain phenotypes fast and accurately. In that chapter, I validated the quality of several morphometric measurements as indicator traits for production traits, which were extracted from 2D and 3D images of gilthead seabream with a novel automated image analysis process. An automated phenotyping process simplifies data collection on large numbers of fish and selecting on indicator traits may boost genetic progress because it allows using Mendelian sampling variance for carcass traits. Chapter 3 addresses the challenge of controlling the costs of integrated breeding programs. In this chapter, I proposed a novel genomic selection design that eliminates the need to genotype a reference population. This novel design uses indicator traits' records of selection candidates as performance data and reduces the cost of genotyping in comparison to traditional genomic selection designs. Chapter 4 addresses the challenge of optimizing logistics and costs related to phenotyping. In this chapter, I investigated the balance between the selection response in harvest weight and a difficult-to-measure trait in a two-stage genomic selection design where genotyping effort is fixed but phenotyping effort for difficult-to-measure trait is not. Depending on the priorities of fish farms, the proposed schemes can balance the phenotyping effort for desired selection responses in harvest weight and a difficult-to-measure trait. In Chapter 5, I estimated GxE for production traits and organ weights between distinct production locations of gilthead seabream. The degree of GxE enables evaluation in terms of expected performance of genetically improved material in distinct locations, and therefore, assists to determine if separate breeding programs are necessary for those environments or if data from those environments need to be combined in a selection index in a single breeding program. If the interest is in reaching a balanced performance across locations, several methods are available to group different types of

environments in a breeding program (Chenu, 2015). A balanced performance across locations might be of interest in integrated breeding programs if a fish farm uses multiple distinct locations to produce fish.

6.2 The need for integrated breeding programs

One significant aspect of integrated breeding programs is that they provide tailor-made genetic improvement because the breeding goal traits are measured on candidates raised in the commercial production environment. Because the target traits are measured in the target environment, GxE is not expected to be an issue in an integrated breeding program. GxE occurs as scaling or reranking of genotypes under distinct environmental factors. In aquaculture, environmental factors can be categorized as natural environmental conditions and management practices. Natural environmental conditions are those related to water conditions that cannot be controlled. However, management practices can be controlled by farmers.

In Chapter 5, I showed that natural water conditions vary between seabream farming locations in the Mediterranean that included temperature profile, oxygen concentrations, and salinity levels. In this chapter, I detected strong GxE for production traits including growth and fillet yield, which is most likely because of differences in those natural water conditions. Strong GxE between farming locations shows that there is a need for tailored-to-the-location genetic improvement.

To show that management practices also differ between farms, I would like to report a summary of a multidisciplinary survey that was conducted in 2018 within the MedAID (Horizon 2020) project (Cidad et al., 2018). The survey was designed by Institute of Agrifood Research and Technology (IRTA), and the questions aimed at collecting information about production data and management practices of seabream farms in the Mediterranean (Muniesa et al., 2020). For the period of January 2015 - December 2017, 15 seabream farms from nine countries provided answers for the survey. Answers demonstrate the variation in

management practices. For example, the maximum rearing density in sea cages varied from as low as 8 kg/m^3 to as high as 70 kg/m^3. Rearing density has been reported to vary in other species as well and can cause GxE (Dupont-Nivet et al., 2008). The feed composition varied between 38 – 56% for protein content, 15 – 21% for fat content, and 15 – 96 MJ/kg for digestible energy. The feed type also differed among the farms that included extruded and non-extruded feeds. Genotype by diet interaction is a specific type of GxE and known to occur in aquaculture (Francis et al., 2021; Le Boucher et al., 2013). Therefore, tailor-made genetic improvement is necessary in the presence of different management practices.

For genetic improvement of fish species whose farming conditions differ vastly from one farm to another as in gilthead seabream, integrated breeding programs may be preferred over central-nucleus approach. There are several reasons for this. First, the genetic improvement is tailor-made in integrated approach because the breeding goal traits are measured in the production environment. This diminishes the risk of suboptimal performance due to GxE and provides the strongest buffer against evolving water conditions caused by climate change. Second, the maintenance of selection candidates does not involve extra costs for the fish farm because they are chosen from the production population. Third, the production population is produced from the latest generation of selected broodstock and therefore, genetic improvement is realized without any time lag. Fourth, the benefit of genetic improvement is realized through increased output of fish farm and therefore, breeding program is an additional step in and integral part of production. Fifth, each element of the breeding program can be shaped according to specific needs of the fish farm. In addition, I find the integrated approach as the most feasible option for genetic improvement of fish species that are produced at small quantities in limited areas. For these species, an integrated approach may boost income through increased production at little extra cost.

6.3 Improvement of integrated breeding programs

In this chapter I will describe how the results of my thesis can contribute to improvement of integrated breeding programs in terms of optimizing efforts for logistics, reducing cost, and decreasing complexity.

6.3.1 Logistics and cost

6.3.1.1 Phenotyping

Phenotyping comprises the most time-consuming and laborious part of fish breeding programs, and therefore in certain cases, it may be a bottleneck for genetic improvement (Fu and Yuna, 2022) because phenotyping efforts are restricted by available resources. Phenotypes are quantification of traits, and their measurement accuracy directly affects the quality of genetic analyses. In aquaculture, there are various traits of commercial and societal interest. Traits of commercial interest include harvest weight (Ponzoni et al., 2008), fillet yield (Fraslin et al., 2018), and feed conversion ratio (Besson et al., 2020) among other traits. These traits can be quantified on a continuous scale. In addition, there are traits quantified on a limited scale, for example survival measured as dead or alive (0 or 1) (Dinh Pham et al., 2021), and color of the fillet as 10 categories (as consecutive numbers from 25 to 34) (Gümüş et al., 2023). An example trait of societal interest is animal welfare, which is quantified as a combination of behavioral traits, disease resistance, and the severity of deformities (Grimsrud et al., 2013). Traits can also be classified whether they are difficult-to-measure.

Difficult-to-measure traits include the ones that require damaging the fish to be quantified and that can be measured on live fish but are laborious or expensive to quantify. Examples of traits that require damaging are slaughter traits including fillet yield, and meat quality traits including texture and fatty acid profiling. Disease resistance traits are quantified in challenge tests that damage the fish including skin lesions and death (Van Doan et al., 2019). Improvement of difficult-to-measure traits add value to the farmed fish and increases the profit of farms and reduces environmental impact in case of improved feed conversion

ratio (Besson et al., 2020), prevents economic losses in case of increased survival (Chiu et al., 2010), improves the value of the harvested product in case of improved fillet yield (Gjedrem, 2008), and increases the sensory and nutritional characteristics, and potentially the price of fillet in case of improved meat quality characteristics (Alfnes et al., 2006; Horn et al., 2020; Johnston et al., 2000). Most central-nucleus breeding programs for fish species include difficult-to-measure traits in their breeding goals (Chavanne et al., 2016).

In recent years, novel phenotyping methods were developed using image analysis that facilitate quantification of traits. These methods commonly involve measurement of morphometric traits from images of fish, either manually or automated. Manual measurements are done by hand marking the phenotype of interest using computer software, as performed by Cardoso et al. (2021), who have manually quantified trunk, fillet, head, and tail areas from Nile tilapia images. More sophisticated image analysis methods involve automated measurements, as performed by Navarro et al. (2016), who automated the quantification of 27 morphometric traits of four fish species. Recently, Liao et al., (2021) have used 3-dimensional (3D) images to automatically extract volume related 3D phenotypes of Tibetan endemic fish species. Automated image analysis is highly advantageous in comparison to classical manual phenotyping and can yield several phenotypes in seconds with little labor (Navarro et al., 2016). However, images of fish yield only morphometric measurements that are related to the shape of the fish. With image analysis, it is not possible to obtain measurements of difficult-to-measure traits unless the fish is slaughtered for that purpose, as performed by Stien et al. (2007), who analyzed salmon fillet images for color.

Difficult-to-measure traits cannot be quantified directly from images unless the fish is slaughtered; however, morphometric traits from images can serve as indicator traits for difficult-to-measure traits. In Chapter 2, I estimated genetic parameters of morphometric traits that were quantified with image analysis,

including their genetic correlations to production traits. The results of Chapter 2 are promising to support the idea that automated phenotyping with image analysis has the potential to replace traditional manual phenotyping in fish breeding programs. In integrated breeding programs, image analysis is preferable to traditional phenotyping methods due to several reasons. The first reason is that image analysis can be automated. To perform automated phenotyping, software should be created to process the images and then each fish needs to be photographed in the breeding program. The second reason is that image analysis is cost-effective. After investing in imaging equipment, automated phenotyping saves costly labor and valuable time compared to slaughtering fish to obtain phenotypes for difficult-to-measure traits. Image analysis saves time in phenotyping for harvest weight as well. Chapter 2 shows that there are morphometric traits with a perfect correlation to harvest weight. This means that harvest weight can indirectly be selected for without any loss in selection response. The third reason is that automated image analysis is fast. Navarro et al. (2016) reports that each 2-dimensional (2D) image takes ~10 seconds to be processed to yield 27 morphometric measurements. Liao et al., (2021) reports that each 3D image takes ~50 seconds to be processed to yield volume measurements. In Chapter 2, the image acquisition time was ~10 seconds. Each 2D and 3D image was processed in ~0.3 seconds to yield 40 morphometric measurements. The fourth reason is that quantification with image analysis is standardized with software and the phenotypes are more accurate and reproducible compared to manual phenotyping. Manual measurements are prone to human error. For example, Rutten et al. (2004) reported that the filleter has considerable influence on fillet yield. Reduced measurement errors are expected to yield higher heritabilities. For example, Elalfy et al. (2021) reported that automated measurements of morphometric traits yielded on average 24% higher heritabilities than manual measurements. Improved heritabilities contribute to higher selection responses (Falconer and Mackay, 1996). Overall, image analysis contributes to solving the phenotyping challenge in integrated

breeding programs. There are numerous reasons to prefer automated phenotyping over traditional phenotyping methods, especially for difficult-to-measure traits.

Genotype by environment interactions may affect the phenotyping strategies in integrated breeding programs. Integrated approach provides tailored-to-the-location genetic improvement; however, fish farms may have distinct production locations in terms of geographical location or water conditions. If there is strong GxE, a strategy is necessary to reach a balanced performance in these locations rather than a high performance in a single location. Combining phenotypes measured in these locations in a selection index may balance the genetic improvement across production environments (Chu et al., 2018). In Chapter 5, I documented that there is strong GxE for main production related traits of seabream in different parts of the Mediterranean. Therefore, fish farms that produce in distinct environments should be aware that GxE can reduce the effectiveness of integrated breeding programs if the genetic improvement is not fully realized in all the locations. An interesting observation in Chapter 5 is that the genetic variance of production traits is considerably different in distinct environments. Genetic variances and consequently, heritabilities influence the selection response. Therefore, farms may consider collecting selection candidates in the location where the genetic variances are the highest. Phenotyping strategies in integrated breeding programs should be based on the degree of GxE between production environments, the genetic parameters of the traits, and production volumes in those environments.

Morphometric traits quantified with image analysis may display less GxE across environments than manually quantified traits. Elalfy et al. (2021) have reported genetic parameters of several manually and digitally quantified traits of seabream between two distinct environments including GxE estimates. They have reported that manually measured harvest weight and digitally measured lateral length in seabream are genetically the same trait since they have a

perfect genetic correlation. In addition, while the magnitude of GxE was strong for weight (0.46 ± 0.24), it was much weaker for lateral length (0.73 ± 0.17). Furthermore, Hasan et al. (2022) have performed a similar experiment in black tiger shrimp and obtained comparable results. Similarly, they have reported that manually measured weight and digitally measured length are genetically the same trait. The magnitudes of GxE across several tanks were lower for length than for weight, although the differences in magnitudes were slight. Therefore, these results may indicate that automated measurements with image analysis are less prone to GxE.

6.3.1.2 Genotyping

Genotyping is the process of determining the genetic makeup of individuals, which is a necessity for genomic selection. Genomic selection is a relatively recent technology in animal breeding that relies on estimating the effects of markers to obtain breeding values of individuals (Meuwissen et al., 2001). This technology has several advantages over traditional pedigree-based selection methods, most importantly capturing the Mendelian sampling effect that allows highly accurate and unique assignment of estimated breeding values for each individual. Despite the advantages, the implementation of genomic selection is marginal and experimental in fish breeding programs, except for a few high-value fish species (Boudry et al., 2021). Financial resources play a restricting role in adopting genomic selection because genotyping costs can be inhibitive. Therefore, research has focused on reducing genotyping costs including lowering the density of SNP panels (Kriaridou et al., 2020), imputation of SNPs (Deng et al., 2022), and not performing sib testing every generation (Sonesson and Meuwissen, 2009).

Cost is a major factor in preferring integrated breeding programs over central-nucleus approach. Despite that, I suggest utilizing genomic selection in integrated breeding programs because of the advantages this technology offers. In integrated breeding programs, anonymous fish in production environment are

the selection candidates and they need to be identified for genetic analyses. It is possible to determine the exact relationship between anonymous animals with genotyping technology. Although low-density SNP arrays can be used to determine the pedigree only, in my opinion, genotyping with high-density SNP arrays are more advantageous despite the costs because genomic selection accelerates genetic improvement. SNP arrays have been developed for fish species in high demand including Atlantic salmon (Houston et al., 2014), rainbow trout (Palti et al., 2015), and tilapia (Joshi et al., 2018). The integrated approach is well suited for fish species that are farmed at a relatively small scale, for which typically no commercial SNP arrays are available. Therefore, the entire genome of these species must be sequenced. However, whole genome sequencing is more costly than genotyping with SNP arrays (D'Agaro et al., 2021), which is a hindrance for integrated breeding programs.

Chapter 3 contributes to solving the challenge of cost reduction for genotyping to promote widespread implementation of genomic selection. In Chapter 3, I established a novel genomic selection design in which there is no reference population; therefore, no associated genotyping costs. Genotyping reference populations in fish breeding can incur high costs because data of large amounts of siblings may be necessary to fully benefit from genomic selection (Boudry et al., 2021). The novel genomic selection design uses predicted phenotypes of selection candidates as performance data. The phenotypic prediction equations are established on a population that is imaged prior to slaughter, and then used on imaged selection candidates to predict phenotypes of slaughter traits. This novel design has elevated benefits for integrated breeding programs because both the slaughtered population and selection candidates are sampled from the same production environment.

I expect the phenotypic predictions to be more accurate in this case because exposure to similar environmental conditions is likely to increase the physiological uniformity of a population. The phenotypic prediction models may

contain several morphometric measurements as predictors, as shown by Sang et al. (2009) and Vandeputte et al. (2017). The accuracy of these models is sensitive to changes in the variation of and the relationship between morphometric measurements. The relationship, specifically, the proportions between morphometric measurements is affected by the size of the fish (Bosworth et al., 1998). Therefore, I expect the highest prediction accuracy if the slaughtered fish and selection candidates have the same mean and variation in weight. This holds true in integrated breeding programs; however, may not hold true in nucleus approaches where selection candidates are isolated in a bio-secure environment and typically larger in size than production (slaughtered) animals. Together with size, differences in environmental conditions may also affect the proportions between morphometric measurements. In Chapter 5, I documented that closely related fish reared in different environments to the same age can have different mean and variation in physiological measures, as well as a different relationship between them in terms of genetic and phenotypic correlations. The relationship between morphometric measurements as physiological traits can be affected by environment, as shown by León-Bernabeu et al. (2021) who detected GxE for a morphometric measurement between different production environments. Contrarily, there are studies that reported no GxE for morphometric traits (Elalfy et al., 2021; Freitas et al., 2021). Therefore, the relationship between morphometric measurements and the accuracy of phenotypic prediction models are influenced by the degree of similarity between environments, the age and size of fish, and the species in question among other factors.

6.3.2 Complexity

6.3.2.1 Breeding goals and selection criteria in combination with image analysis

Breeding goals are directions for improvement in breeding programs in terms of performance and defining them is the first step of systematic genetic

improvement (Groen, 2000). Increased growth rate is typically an initial direction and a common breeding goal in fish breeding programs (de Oliveira et al., 2016; Gjedrem, 2010; Sui et al., 2019). Advanced breeding programs may include several other traits in their breeding goals in addition to growth rate (Chavanne et al., 2016). To the best of my knowledge, integrated fish breeding programs currently aim to improve growth rate only. However, additional traits should be included in the breeding goals of integrated breeding programs to have more efficient, more resilient, and more resistant production animals.

Selection criteria are the specific traits on which the selection is performed. In Chapter 5, I showed that harvest weight is an almost perfect criteria to improve growth of seabream because of close-to-unity genetic correlation between these traits. Weight is one of the easiest traits to quantify in fish breeding programs. Despite that, weighing takes considerable amount of effort because each fish needs to be taken out of water and handled individually. In addition to the labor, this process is stressful and potentially damaging for the fish in case of improper handling (Wedemeyer, 1996).

Automated image analysis offers an easier way to quantify traits. In Chapter 3, I showed that harvest weight can be predicted perfectly using automated morphometric measurements, which were extracted with image analysis methods built in Chapter 2. The measurements were automated in the sense that after a fish image is available, the software processes the images and yields the predefined measurements. In Chapter 2, we imaged the fish outside of the water, which stresses the fish. To prevent that, underwater cameras can be used to image the fish within the production system, such as in sea cages (Saberioon et al., 2017), which can lead to fully automated phenotyping. However, there are a few disadvantages associated with underwater imaging. First, it requires special saltwater-resistant equipment. In addition to the cost, these cameras are harder to maintain than regular cameras. The second disadvantage is that the quality of the underwater images may not be uniform. The illumination is likely to be

suboptimal, especially when the water is not clear (Zhou et al., 2017a). In addition, the position of the fish in each image is likely to be different because they swim freely (Zhou et al., 2017b). The variations in image quality in terms of illumination and positioning may negatively influence the measurement accuracy (Wright et al., 2020). Deep learning methods may suit more for the analysis of underwater images because deep learning models can be trained to be robust to alterations in image quality (Akhtar and Mian, 2018). The third disadvantage is that the identity of the fish should be known prior to imaging. If the fish is taken out of water, tagging and imaging can be performed together, but this is not possible underwater. This creates limitations in terms of the number and location of the selection candidates. For example, collecting selection candidates from several sea cages may not be possible because of the practical difficulties of finding a limited number of tagged fish among all the production animals. A solution to this obstacle may be rearing the selection candidates in a smaller cage where all the fish is tagged. This also reduces the logistics workload, as it is much easier to collect selection candidates from a smaller cage rather than full-size production cages. Another solution for underwater identification is with deep learning applications, which identify individual fish from images or videos without physical tagging. However, the identification accuracy of these applications may hinder their use in integrated breeding programs (Yang et al., 2021). In addition, the fish still must be taken out of water for DNA collection if genomic selection is intended.

Phenotyping with image analysis has elevated benefits if slaughter traits are in the breeding goals. In Chapter 2, I showed that indirect selection for fillet yield is highly efficient with morphometric traits. In Chapter 3, I showed that indirect selection with image analysis has the potential to boost the genetic improvement compared to direct selection with manual phenotyping because of high heritabilities of certain morphometric traits and their almost unity genetic correlation to fillet yield. The genetic parameters of the same traits can change

greatly between populations (Gjerde et al., 2012; Nguyen et al., 2010; Rutten et al., 2005). To obtain population specific genetic parameters, analyses must be performed with population specific data, which requires the slaughter of fish to obtain records of slaughter traits. However, filleting is laborious and time consuming. On the other hand, the slaughter data can be obtained routinely if farms have their own slaughterhouses. In addition, processing farm's own harvest creates extra value for improved fillet yield. To obtain financial benefits through improved fillet yield, farmers should be paid on the amount of fillet rather than on total fish weight. Farmers of high-value fish species in certain countries are paid on the amount of fillet. Such as, Sang et al. (2009) reported that the catfish processing companies in Vietnam pay the farmers based on a combination of harvest weight and fillet yield. Similarly, Nguyen et al. (2010) reported that in major tilapia-producing countries, the payment is shifting from harvest weight to fillet weight. However, it is not likely that processing companies will pay for the fillet of species that are farmed at a small scale. Therefore, fish farms can add more value to their harvest if they have slaughterhouses, in which case they can sell fillet rather than the whole fish.

6.3.2.2 Multi-stage genomic selection for fillet yield in combination with image analysis

In Chapter 4, I focused on optimizing phenotyping efforts for slaughter traits, *e.g.,* for fillet yield, in genomic selection programs. The main implication of Chapter 4 for integrated breeding programs is that breeding programs must choose if they desire a higher response in harvest weight or in fillet yield because a higher response in one trait leads a lower response in the other under genotyping constraints. Logically, this choice is based on the income generated by genetic improvement in traits. In case the main income of a farm is generated by selling whole fish, producing biomass is more important than producing more fillet. However, in case the main income of a farm is generated by selling fillets, the fillet yield might be more important than the biomass produced. In Chapter 4, I

simulated a two-stage selection scheme, which is used widely in fish breeding programs (Cao et al., 2012; Martinez et al., 2006; Sae-Lim et al., 2013). An advantage of two-stage selection is that it saves time and resources because weight is typically the only trait measured in the first stage of selection. Reduction of labor is essential for the implementation of integrated breeding programs. Therefore, the selection scheme in Chapter 4 can be combined with image analysis for a much simpler implementation of multi-stage selection. In the first stage, the fish can be selected on an indicator trait such as *total volume* that is perfectly correlated to harvest weight. Another possibility with image analysis is to mass select on fillet yield through selection on indicator traits, such as *trunk volume*. Likewise in the second stage, indicator traits can be used instead of manual measurements. Using image analysis may reduce the number of fish that need to be slaughtered in genomic selection. In Chapter 3, I showed that the accuracy of genomic estimated breeding values increases with more performance data even if the performances are predicted. Therefore, I expect that increasing the number of fish with predicted performance data will contribute to larger increases in the accuracy of breeding values than increasing the number of slaughtered fish that are used to build phenotypic prediction equations.

6.4 Further improvement of integrated breeding programs: Real-time selection

So far, I have focused mostly on improving phenotyping strategies in integrated breeding programs. Now, I will outline how breeding programs can further be improved by combining image analysis with high-throughput algorithms to perform real-time selection. In breeding programs, the steps of data collection, statistical analysis, and selection of animals are discrete and consecutive (Shepherd and Kinghorn, 1993). Ideally, these discrete steps can be combined into a single step for more efficient use of time and resources. I call this process

"real-time selection", which allows to keep or discard a selection candidate as soon as it is phenotyped.

For a preview of how real-time selection might work, I simulated a population of 1,000 fish with equal numbers of males and females with a single trait. The trait was quantified on selection candidates and the heritability varied between 0.1 – 0.7. The simulated real-time selection process phenotyped the selection candidates one by one in a random order. When a new phenotype was available, the process updated the running mean and standard deviation of the selection criteria. The selection was either on phenotypes or on estimated breeding values (EBV). In case of mass selection, the running mean was calculated as $\mu R_n = (P_1 + P_2 + P_3 + \cdots + P_n)/n$ where *P* is a phenotype and *n* is the number of fish phenotyped including the last one, and a running standard deviation as $\sigma R_n = \sqrt{\sum(P_i - \mu R_n)^2/(n-1)}$ where P_i is the phenotype of the last measured fish. In case of pedigree selection, the running mean was calculated as $\mu R_n = (EBV_1 + EBV_2 + EBV_3 + \cdots + EBV_n)/n$ and a running standard deviation as $\sigma R_n = \sqrt{\sum(EBV_i - \mu R_n)^2/(n-1)}$. The process kept or discarded a selection candidate by comparing its phenotype or estimated breeding value to a threshold, which was the quantile of $(N_{no\_parents} - N_{selected})/(N-(n-1))$ in a distribution of $N(\mu Rn, \sigma^2 Rn)$ where $N_{no\_parents}$ was the number of required parents, $N_{selected}$ was the number selection candidates kept until the last phenotyping, *N* was the total number of selection candidates, and *n* was the number of animals phenotyped including the last one. The process kept a selection candidate to be a parent if its phenotype or estimated breeding value was higher than the threshold.

I calculated the accuracy of the real-time selection process by comparing the resulting selection response to a gold-standard reference response. In the case of mass selection, reference response was obtained by first phenotyping all selection candidates and determining the ones with the highest phenotypes. In the case of pedigree selection, reference response was obtained by estimating

the breeding values after all phenotypes were available and determining the ones with the highest EBVs. The selected proportion was 15% in all the scenarios. For mass selection, the real-time process yielded 99.5% of the gold-standard reference selection response independent of heritability. For pedigree-selection, the real-time process yielded 81% of the gold-standard reference selection response for a heritability of 0.1, which increased linearly to 98% until the heritability of 0.7.

The high accuracy of real-time selection is promising for its implementation in integrated breeding programs. If real-time selection technology is combined with image analysis technology, the entire selection process can be fully automated. For example, the mass selection stage in Chapter 4 can be performed by real-time selection technology on image-derived measurements for almost-fully automatic selection. However, the implementation of real-time selection process in genomic selection is more challenging because it requires rapid handling of large genomic data. Therefore, optimization of the real-time selection technology is necessary for quick handling of genomic data.

6.5 Concluding remarks

I believe that expansion in aquaculture production is essential to meet the growing demand for food in the world. The sustainable development of aquaculture depends on several factors including advances in aquaculture engineering (Lekang, 2020), feed technology (Paolucci, 2023) and not the least governmental policies (Bostock et al., 2016). Among all factors, genetic improvement plays a pivotal role because the genetic makeup of production animals greatly influences their performance (Næve et al., 2022). Therefore, it is essential to use genetically improved stock in aquaculture. Despite the advantages of genetic improvement, a great majority of farmed fish species are not genetically improved. However, all fish species deserve to have a breeding program to become high yielding and efficient under production conditions. Genetic improvement of lesser-known fish species will benefit their farmers with

increased income and improve the species diversity in aquaculture production. Society will also benefit from an increased choice of fish to purchase at the market and possibly reduced prices.

The research I conducted in this thesis contributes highly to improved implementation of integrated breeding programs in terms of optimizing efforts and reducing complexity. Image analysis is the master key to unlock the full potential of integrated breeding programs because automated phenotyping is in the center of all pieces for improvement. The novel genomic selection design depends on automated phenotyping and is fundamental to make the breeding program affordable. Two-stage selection is another essential component for reducing the complexity of the breeding program. GxE studies will assist the breeders on data collection strategies within integrated breeding programs and the optimal environments to keep the selection candidates. Further research on innovative technologies will facilitate the implementation of integrated breeding programs, which is fundamental for selective breeding of countless farmed fish species.

References

Akhtar, N., Mian, A., 2018. Threat of adversarial attacks on deep learning in computer vision: A survey. Ieee Access 6, 14410–14430.

Alfnes, F., Guttormsen, A.G., Steine, G., Kolstad, K., 2006. Consumers' Willingness to Pay for the Color of Salmon: A Choice Experiment with Real Economic Incentives. Am J Agric Econ 88, 1050–1061. https://doi.org/10.1111/j.1467-8276.2006.00915.x

Aslam, M.L., Carraro, R., Sonesson, A.K., Meuwissen, T., Tsigenopoulos, C.S., Rigos, G., Bargelloni, L., Tzokas, K., 2020. Genetic Variation, GWAS and Accuracy of Prediction for Host Resistance to Sparicotyle chrysophrii in Farmed Gilthead Sea Bream (Sparus aurata). Front Genet.

Bentsen, H.B., Gjerde, B., Eknath, A.E., de Vera, M.S.P., Velasco, R.R., Danting, J.C., Dionisio, E.E., Longalong, F.M., Reyes, R.A., Abella, T.A., Tayamen, M.M., Ponzoni, R.W., 2017. Genetic improvement of farmed tilapias: Response to five generations of selection for increased body weight at harvest in Oreochromis niloticus and the further impact of the project. Aquaculture 468, 206–217. https://doi.org/10.1016/J.AQUACULTURE.2016.10.018

Besson, M., Komen, H., Rose, G., Vandeputte, M., 2020. The genetic correlation between feed conversion ratio and growth rate affects the design of a breeding program for more sustainable fish production. Genetics Selection Evolution 52, 5. https://doi.org/10.1186/s12711-020-0524-0

Blonk, R.J.W., Komen, J., Tenghe, A., Kamstra, A., van Arendonk, J.A.M., 2010. Heritability of shape in common sole, Solea solea, estimated from image analysis data. Aquaculture 307, 6–11. https://doi.org/10.1016/j.aquaculture.2010.06.025

Bolormaa, S., Gore, K., van der Werf, J.H.J., Hayes, B.J., Daetwyler, H.D., 2015. Design of a low-density SNP chip for the main Australian sheep breeds and its effect on imputation and genomic prediction accuracy. Anim Genet 46, 544–556. https://doi.org/10.1111/age.12340

Bostock, J., Lane, A., Hough, C., Yamamoto, K., 2016. An assessment of the economic contribution of EU aquaculture production and the influence of policies for its sustainable development. Aquaculture International 24, 699–733.

Bosworth, B.G., Libey, G.S., Notter, D.R., 1998. Relationships Among Total Weight, Body Shape, Visceral Components, and Fillet Traits in Palmetto Bass (Striped Bass Female Morone saxatilis× White Bass Male M. chrysops) and Paradise Bass (Striped Bass Female M. saxatilis× Yellow Bass Male M. mississippie. J World Aquac Soc 29, 40–50. https://doi.org/10.1111/j.1749-7345.1998.tb00298.x

Boudry, P., Allal, F., Aslam, M.L., Bargelloni, L., Bean, T.P., Brard-Fudulea, S., Brieuc, M.S.O., Calboli, F.C.F., Gilbey, J., Haffray, P., Lamy, J.-B., Morvezen, R., Purcell, C., Prodöhl, P.A., Vandeputte, M., Waldbieser, G.C., Sonesson, A.K., Houston, R.D., 2021. Current status and potential of genomic selection to improve selective breeding in the main aquaculture species of International Council for the Exploration of the Sea (ICES) member countries. Aquac Rep 20, 100700. https://doi.org/10.1016/j.aqrep.2021.100700

Bourdon, R.M., 2000. Understanding animal breeding. Prentice Hall Upper Saddle River, NJ.

Calus, M.P.L., Vandenplas, J., 2019. Calc_grm—a program to compute pedigree, genomic, and combined relationship matrices. Animal Breeding and Genomics Centre, Wageningen UR Livestock Research.

Camara, M.D., Symonds, J.E., 2014. Genetic improvement of New Zealand aquaculture species: programmes, progress and prospects. N Z J Mar Freshwater Res 48, 466–491. https://doi.org/10.1080/00288330.2014.932291

Cao, X.J., Wang, H.P., Yao, H., O'Bryant, P., Rapp, D., Wang, W.M., MacDonald, R., 2012. Evaluation of 1-stage and 2-stage selection in yellow perch I: Genetic and phenotypic parameters for body weight of F1 fish reared in ponds using microsatellite parentage assignment1. J Anim Sci 90, 27–36. https://doi.org/10.2527/jas.2011-3902

Cardoso, A.J. da S., Oliveira, C.A.L. de, Campos, E.C., Ribeiro, R.P., Assis, G.J. de F., Silva, F.F. e, 2021. Estimation of genetic parameters for body areas in Nile tilapia measured by digital image analysis. Journal of Animal Breeding and Genetics n/a. https://doi.org/10.1111/jbg.12551

Chavanne, H., Janssen, K., Hofherr, J., Contini, F., Haffray, P., Komen, H., Nielsen, E.E., Bargelloni, L., 2016. A comprehensive survey on selective breeding programs and seed market in the European aquaculture fish industry. Aquaculture International 24, 1287–1307. https://doi.org/10.1007/s10499-016-9985-0

Chenu, K., 2015. Chapter 13 - Characterizing the crop environment – nature, significance and applications, in: Sadras, V.O., Calderini, D.F.B.T.-C.P. (Second E. (Eds.), . Academic Press, San Diego, pp. 321–348. https://doi.org/10.1016/B978-0-12-417104-6.00013-3

Chiu, C.-H., Cheng, C.-H., Gua, W.-R., Guu, Y.-K., Cheng, W., 2010. Dietary administration of the probiotic, Saccharomyces cerevisiae P13, enhanced the growth, innate immune responses, and disease resistance of the grouper, Epinephelus coioides. Fish Shellfish Immunol 29, 1053–1059. https://doi.org/10.1016/j.fsi.2010.08.019

Chu, Thinh T, Alemu, S.W., Norberg, E., Sørensen, A.C., Henshall, J., Hawken, R., Jensen, J., 2018. Benefits of testing in both bio-secure and production environments in genomic selection breeding programs for commercial broiler chicken. Genetics Selection Evolution 50, 52. https://doi.org/10.1186/s12711-018-0430-x

Chu, Thinh Tuan, Bastiaansen, J.W.M., Norberg, E., Berg, P., 2018. On farm observations to increase genetic gain in breeding schemes for village poultry production – A simulation study. Acta Agric Scand A Anim Sci 68, 1–10. https://doi.org/10.1080/09064702.2018.1543444

Chu, T.T., Sørensen, A.C., Lund, M.S., Meier, K., Nielsen, T., Su, G., 2020. Phenotypically Selective Genotyping Realizes More Genetic Gains in a Rainbow Trout Breeding Program in the Presence of Genotype-by-Environment Interactions. Front Genet 11, 866. https://doi.org/10.3389/fgene.2020.00866

Cidad, M., Peral, I., Ramos, S., Basurco, B., López-Francos, A., Muniesa, A., Cavallo, M., Perez, J., Aguilera, C., Furones, D., 2018. Assessment of Mediterranean Aquaculture Sustainability.

Cobb-Vantress Inc., 2023. Cobb-Vantress - Generations of Progress: Quality products for all markets and needs [WWW Document]. URL https://www.cobb-vantress.com/en_US/products/ (accessed 3.9.23).

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.

Daetwyler, H.D., Villanueva, B., Woolliams, J.A., 2008. Accuracy of Predicting the Genetic Risk of Disease Using a Genome-Wide Approach. PLoS One 3, e3395.

D'Agaro, E., Favaro, A., Matiussi, S., Gibertoni, P.P., Esposito, S., 2021. Genomic selection in salmonids: new discoveries and future perspectives. Aquaculture International 29, 2259–2289. https://doi.org/10.1007/s10499-021-00747-w

Dagnachew, B., Meuwissen, T., 2019. Accuracy of within-family multi-trait genomic selection models in a sib-based aquaculture breeding scheme. Aquaculture 505, 27–33. https://doi.org/10.1016/j.aquaculture.2019.02.036

de Oliveira, C.A.L., Ribeiro, R.P., Yoshida, G.M., Kunita, N.M., Rizzato, G.S., de Oliveira, S.N., dos Santos, A.I., Nguyen, N.H., 2016. Correlated changes in body shape after five generations of selection to improve growth rate in a breeding program for Nile tilapia Oreochromis niloticus in Brazil. J Appl Genet 57, 487–493. https://doi.org/10.1007/s13353-016-0338-5

de Verdal, H., Vandeputte, M., Mekkawy, W., Chatain, B., Benzie, J.A.H., 2018. Quantifying the genetic parameters of feed efficiency in juvenile Nile tilapia Oreochromis niloticus. BMC Genet 19, 105. https://doi.org/10.1186/s12863-018-0691-y

Deng, T., Zhang, P., Garrick, D., Gao, H., Wang, L., Zhao, F., 2022. Comparison of Genotype Imputation for SNP Array and Low-Coverage Whole-Genome Sequencing Data. Front Genet 12.

Dinh Pham, K., Ødegård, J., Van Nguyen, S., Magnus Gjøen, H., Klemetsdal, G., 2021. Genetic analysis of resistance in Mekong striped catfish (Pangasianodon hypophthalmus) to bacillary necrosis caused by Edwardsiella ictaluri. J Fish Dis 44, 201–210. https://doi.org/10.1111/jfd.13279

Domingos, J.A., Goldsbury, J.A., Bastos Gomes, G., Smith, B.G., Tomlinson, C., Bade, T., Sander, C., Forrester, J., Jerry, D.R., 2021. Genotype by environment interactions of harvest growth traits for barramundi (Lates calcarifer) commercially farmed in marine vs. freshwater conditions. Aquaculture 532, 735989. https://doi.org/10.1016/j.aquaculture.2020.735989

Dowlati, M., de la Guardia, M., Dowlati, M., Mohtasebi, S.S., 2012a. Application of machine-vision techniques to fish-quality assessment. TrAC Trends in Analytical Chemistry 40, 168–179. https://doi.org/10.1016/J.TRAC.2012.07.011

Dowlati, M., de la Guardia, M., Dowlati, M., Mohtasebi, S.S., 2012b. Application of machine-vision techniques to fish-quality assessment. TrAC Trends in Analytical Chemistry 40, 168–179. https://doi.org/10.1016/j.trac.2012.07.011

Dupont-Nivet, M., Vandeputte, M., Vergnet, A., Merdy, O., Haffray, P., Chavanne, H., Chatain, B., 2008. Heritabilities and GxE interactions for growth in the European sea bass (Dicentrarchus labrax L.) using a marker-based pedigree. Aquaculture 275, 81–87. https://doi.org/10.1016/j.aquaculture.2007.12.032

Eckstein, W., Steger, C., 1999. The Halcon vision system: an example for flexible software architecture, in: Proceedings of 3rd Japanese Conference on Practical Applications of Real-Time Image Processing. Citeseer, pp. 18–23.

Elalfy, I., Shin, H.S., Negrín-Báez, D., Navarro, A., Zamorano, M.J., Manchado, M., Afonso, J.M., 2021. Genetic parameters for quality traits by non-invasive methods and their G x E interactions in ocean cages and estuaries on gilthead seabream (Sparus aurata). Aquaculture 537, 736462. https://doi.org/10.1016/j.aquaculture.2021.736462

Falconer, D.S., Mackay, T.F.C., 1996. Introduction to quantitative genetics, Fourth edi. ed. Longman Group, Ltd.

FAO, 2021. Fishery and Aquaculture Statistics. Global aquaculture production 1950-2019 (FishstatJ).

Fernandes, A.F.A., Turra, E.M., de Alvarenga, É.R., Passafaro, T.L., Lopes, F.B., Alves, G.F.O., Singh, V., Rosa, G.J.M., 2020. Deep Learning image segmentation for extraction of fish body measurements and prediction of body weight and carcass traits in Nile tilapia. Comput Electron Agric 170, 105274. https://doi.org/10.1016/j.compag.2020.105274

Fernández, J., Villanueva, B., Toro, M.A., 2021. Optimum mating designs for exploiting dominance in genomic selection schemes for aquaculture species. Genetics Selection Evolution 53, 14. https://doi.org/10.1186/s12711-021-00610-9

Francis, M., Li, C., Sun, Y., Zhou, J., Li, X., Brenna, J.T., Ye, K., 2021. Genome-wide association study of fish oil supplementation on lipid traits in 81,246 individuals reveals new gene-diet interaction loci. PLoS Genet 17, e1009431-.

Fraslin, C., Dupont-Nivet, M., Haffray, P., Bestin, A., Vandeputte, M., 2018. How to genetically increase fillet yield in fish: New insights from simulations based on field data. Aquaculture 486, 175–183. https://doi.org/10.1016/J.AQUACULTURE.2017.12.012

Freitas, M. V, Lira, L.V.G., Ariede, R.B., Agudelo, J.F.G., Oliveira Neto, R.R. de, Borges, C.H.S., Mastrochirico-Filho, V.A., Garcia Neto, B.F., Carvalheiro, R., Hashimoto, D.T., 2021. Genotype by environment interaction and genetic parameters for growth traits in the Neotropical fish pacu (Piaractus mesopotamicus). Aquaculture 530, 735933. https://doi.org/10.1016/j.aquaculture.2020.735933

Fu, G., Yuna, Y., 2022. Phenotyping and phenomics in aquaculture breeding. Aquac Fish 7, 140–146. https://doi.org/10.1016/j.aaf.2021.07.001

García-Ballesteros, S., Fernández, J., Kause, A., Villanueva, B., 2022. Predicted genetic gain for carcass yield in rainbow trout from indirect and genomic selection. Aquaculture 554, 738119. https://doi.org/10.1016/j.aquaculture.2022.738119

Garcia-Celdran, M., Ramis, G., Manchado, M., Estevez, A., Afonso, J.M., Armero, E., 2015. Estimates of heritabilities and genetic correlations of carcass quality traits in a reared gilthead sea bream (Sparus aurata L.) population sourced from three broodstocks along the Spanish coasts. Aquaculture 446, 175–180. https://doi.org/10.1016/j.aquaculture.2015.04.028

García-Ruiz, A., Cole, J.B., VanRaden, P.M., Wiggans, G.R., Ruiz-López, F.J., Van Tassell, C.P., 2016. Changes in genetic selection differentials and generation intervals in US Holstein dairy cattle as a result of genomic selection. Proceedings of the National Academy of Sciences 113, E3995–E4004. https://doi.org/10.1073/pnas.1519061113

Gharbi, K., Matthews, L., Bron, J., Roberts, R., Tinch, A., Stear, M., 2015. The control of sea lice in Atlantic salmon by selective breeding. J R Soc Interface 12, 20150574. https://doi.org/10.1098/rsif.2015.0574

Gilmour, A.R., Gogel, B.J., Cullis, B.R., Welham, Sj., Thompson, R., 2015. ASReml user guide release 4.1 structural specification. Hemel hempstead: VSN international ltd.

Gjedrem, T., 2010. The first family-based breeding program in aquaculture. Rev Aquac 2, 2–15. https://doi.org/10.1111/j.1753-5131.2010.01011.x

Gjedrem, T., 2008. 10 - Improving farmed fish quality by selective breeding, in: Lie, Ø. (Ed.), Improving Farmed Fish Quality and Safety. Woodhead Publishing, pp. 265–274. https://doi.org/10.1533/9781845694920.2.265

Gjedrem, T., 2005. Selection and breeding programs in aquaculture. Springer.

Gjedrem, T., 2000. Genetic improvement of cold-water fish species. Aquac Res 31, 25–33. https://doi.org/10.1046/j.1365-2109.2000.00389.x

Gjedrem, T., Robinson, N., 2014. Advances by selective breeding for aquatic species: a review. Agricultural Sciences 5, 1152.

Gjerde, B., Mengistu, S.B., Ødegård, J., Johansen, H., Altamirano, D.S., 2012. Quantitative genetics of body weight, fillet weight and fillet yield in Nile tilapia (Oreochromis niloticus). Aquaculture 342–343, 117–124. https://doi.org/10.1016/j.aquaculture.2012.02.015

Goddard, M.E., Hayes, B.J., Meuwissen, T.H.E., 2011. Using the genomic relationship matrix to predict the accuracy of genomic selection. Journal of Animal Breeding and Genetics 128, 409–421. https://doi.org/10.1111/j.1439-0388.2011.00964.x

Gonzales, R.C., Woods, R.E., 2018. Digital image processing 4th edition.

Grimsrud, K.M., Nielsen, H.M., Navrud, S., Olesen, I., 2013. Households' willingness-to-pay for improved fish welfare in breeding programs for farmed Atlantic salmon. Aquaculture 372–375, 19–27. https://doi.org/10.1016/j.aquaculture.2012.10.009

Groen, A.F., 2000. Breeding goal definition. Developing breeding Strategies for Lower Input Animal Production Environments 25.

Guan, J., Hu, Y., Wang, M., Wang, W., Kong, J., Luan, S., 2016. Estimating genetic parameters and genotype-by-environment interactions in body traits of turbot in two different rearing environments. Aquaculture 450, 321–327. https://doi.org/10.1016/j.aquaculture.2015.08.014

Gulzari, B., 2017. Genetic parameters and response to selection for body weight, fillet traits, body size traits and survival in Nile Tilapia (Oreochromis niloticus).

Gulzari, B., Komen, H., Nammula, V.R., Bastiaansen, J.W.M., 2022. Genetic parameters and genotype by environment interaction for production traits and organ weights of gilthead seabream (Sparus aurata) reared in sea cages. Aquaculture 548, 737555. https://doi.org/10.1016/j.aquaculture.2021.737555

Gümüş, B., Gümüş, E., Balaban, M.O., 2023. Color of rainbow trout (Oncorhynchus mykiss) fillets by image and sensory analysis, and correlation with SalmoFan numbers. J Food Sci 88, 430–446. https://doi.org/10.1111/1750-3841.16409

Haffray, P., Bugeon, J., Pincent, C., Chapuis, H., Mazeiraud, E., Rossignol, M.-N., Chatain, B., Vandeputte, M., Dupont-Nivet, M., 2012. Negative genetic correlations between production traits and head or bony tissues in large all-female rainbow trout (Oncorhynchus mykiss). Aquaculture 368–369, 145–152. https://doi.org/10.1016/j.aquaculture.2012.09.023

Haffray, P., Bugeon, J., Rivard, Q., Quittet, B., Puyo, S., Allamelou, J.M., Vandeputte, M., Dupont-Nivet, M., 2013. Genetic parameters of in-vivo prediction of carcass, head and fillet yields by internal ultrasound and 2D external imagery in large rainbow trout (Oncorhynchus mykiss). Aquaculture 410–411, 236–244. https://doi.org/10.1016/J.AQUACULTURE.2013.06.016

Hamzah, A., Thoa, N.P., Nguyen, N.H., 2017. Genetic analysis of a red tilapia (Oreochromis spp.) population undergoing three generations of selection for increased body weight at harvest. J Appl Genet 58, 509–519. https://doi.org/10.1007/s13353-017-0411-8

Hao, M., Yu, H., Li, D., 2016. The Measurement of Fish Size by Machine Vision - A Review BT - Computer and Computing Technologies in Agriculture IX, in: Li, D., Li, Z. (Eds.), . Springer International Publishing, Cham, pp. 15–32.

Hasan, M.M., Thomson, P.C., Raadsma, H.W., Khatkar, M.S., 2022. Genetic analysis of digital image derived morphometric traits of black tiger shrimp (Penaeus monodon) by incorporating G× E investigations. Front Genet 13.

Horn, S.S., Aslam, M.L., Difford, G.F., Tsakoniti, K., Karapanagiotis, S., Gulzari, B., Bastiaansen, J.W.M., Peñaloza, C., Houston, R., Ruyter, B., Sonesson, A.K., 2022. Genetic parameters of fillet fatty acids and fat deposition in gilthead seabream (Sparus aurata) using the novel 30 k Medfish SNP array. Aquaculture 556, 738292. https://doi.org/10.1016/j.aquaculture.2022.738292

Horn, S.S., Meuwissen, T.H.E., Moghadam, H., Hillestad, B., Sonesson, A.K., 2020. Accuracy of selection for omega-3 fatty acid content in Atlantic salmon fillets. Aquaculture 519, 734767. https://doi.org/10.1016/j.aquaculture.2019.734767

Horn, S.S., Ruyter, B., Meuwissen, T.H.E., Hillestad, B., Sonesson, A.K., 2018. Genetic effects of fatty acid composition in muscle of Atlantic salmon. Genetics Selection Evolution 50, 23. https://doi.org/10.1186/s12711-018-0394-x

Houston, R.D., Taggart, J.B., Cézard, T., Bekaert, M., Lowe, N.R., Downing, A., Talbot, R., Bishop, S.C., Archibald, A.L., Bron, J.E., 2014. Development and validation of a high density SNP genotyping array for Atlantic salmon (Salmo salar). BMC Genomics 15, 1–13.

Janssen, Kasper, Berentsen, P., Besson, M., Komen, H., 2017. Derivation of economic values for production traits in aquaculture species. Genetics Selection Evolution 49, 5. https://doi.org/10.1186/s12711-016-0278-x

Janssen, K, Chavanne, H., Berentsen, P., Komen, H., 2017. Impact of selective breeding on European aquaculture. Aquaculture 472, 8–16. https://doi.org/10.1016/j.aquaculture.2016.03.012

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.

Johnston, I.A., Alderson, R., Sandham, C., Dingwall, A., Mitchell, D., Selkirk, C., Nickell, D., Baker, R., Robertson, B., Whyte, D., Springate, J., 2000. Muscle fibre density in relation to the colour and texture of smoked Atlantic salmon (Salmo salar L.). Aquaculture 189, 335–349. https://doi.org/10.1016/S0044-8486(00)00373-2

Joshi, R., Almeida, D.B., da Costa, A.R., Skaarud, A., de Pádua Pereira, U., Knutsen, T.M., Moen, T., Alvarez, A.T., 2021. Genomic selection for resistance to Francisellosis in commercial Nile tilapia population: Genetic and genomic parameters, correlation with growth rate and predictive ability. Aquaculture 537, 736515. https://doi.org/10.1016/j.aquaculture.2021.736515

Joshi, R., Árnyasi, M., Lien, S., Gjøen, H.M., Alvarez, A.T., Kent, M., 2018. Development and validation of 58K SNP-array and high-density linkage map in Nile tilapia (O. niloticus). Front Genet 9, 472.

Kause, A., Kiessling, A., Martin, S.A.M., Houlihan, D., Ruohonen, K., 2016. Genetic improvement of feed conversion ratio via indirect selection against lipid deposition in farmed rainbow trout (Oncorhynchus mykiss Walbaum). British Journal of Nutrition 116, 1656–1665.

Kause, A., Nousiainen, A., Koskinen, H., 2022. Improvement in feed efficiency and reduction in nutrient loading from rainbow trout farms: the role of selective breeding. J Anim Sci 100, skac214. https://doi.org/10.1093/jas/skac214

Khaw, H.L., Ponzoni, R.W., Hamzah, A., Abu-Bakar, K.R., Bijma, P., 2012. Genotype by production environment interaction in the GIFT strain of Nile tilapia (Oreochromis niloticus). Aquaculture 326–329, 53–60. https://doi.org/10.1016/j.aquaculture.2011.11.016

Kriaridou, C., Tsairidou, S., Houston, R.D., Robledo, D., 2020. Genomic Prediction Using Low Density Marker Panels in Aquaculture: Performance Across Species, Traits, and Genotyping Platforms. Front Genet.

Le Boucher, R., Vandeputte, M., Dupont-Nivet, M., Quillet, E., Ruelle, F., Vergnet, A., Kaushik, S., Allamellou, J.M., Médale, F., Chatain, B., 2013. Genotype by diet interactions in European sea bass (Dicentrarchus labrax L.): Nutritional challenge with totally plant-based diets1. J Anim Sci 91, 44–56. https://doi.org/10.2527/jas.2012-5311

Lee, J.J., 2023. Cobb-Vantress - Our Story [WWW Document]. URL https://www.cobb-vantress.com/en_US/who-we-are/ (accessed 3.9.23).

Lekang, O.-I., 2020. Aquaculture engineering. John Wiley & Sons.

León-Bernabeu, S., Shin, H.S., Lorenzo-Felipe, Á., García-Pérez, C., Berbel, C., Elalfy, I.S., Armero, E., Pérez-Sánchez, J., Arizcun, M., Zamorano, M.J., Manchado, M., Afonso, J.M., 2021. Genetic parameter estimations of new traits of morphological quality on gilthead seabream (Sparus aurata) by using IMAFISH_ML software. Aquac Rep 21, 100883. https://doi.org/10.1016/j.aqrep.2021.100883

Liao, Y.-H., Zhou, C.-W., Liu, W.-Z., Jin, J.-Y., Li, D.-Y., Liu, F., Fan, D.-D., Zou, Y., Mu, Z.-B., Shen, J., 2021. 3DPhenoFish: Application for two-and three-dimensional fish morphological phenotype extraction from point cloud analysis. Zool Res 42, 492.

Loukovitis, D., Sarropoulou, E., Tsigenopoulos, C.S., Batargias, C., Magoulas, A., Apostolidis, A.P., Chatziplis, D., Kotoulas, G., 2011. Quantitative trait loci involved in sex determination and body growth in the gilthead sea bream (Sparus aurata L.) through targeted genome scan. PLoS One 6, e16599.

Martinez, V., Kause, A., Mäntysaari, E., Mäki-Tanila, A., 2006. The use of alternative breeding schemes to enhance genetic improvement in rainbow trout: II. Two-stage selection. Aquaculture 254, 195–202. https://doi.org/10.1016/j.aquaculture.2005.11.011

Mayer, P., Estruch, V.D., Jover, M., 2012. A two-stage growth model for gilthead sea bream (Sparus aurata) based on the thermal growth coefficient. Aquaculture 358–359, 6–13. https://doi.org/10.1016/j.aquaculture.2012.06.016

Megahed, M.E., 2020. Genetic selection for improved disease resistance in aquaculture with special reference to shrimp and tilapia breeding programs in Egypt. Journal of Applied Aquaculture 32, 291–340. https://doi.org/10.1080/10454438.2019.1697783

Mengistu, S.B., Mulder, H.A., Benzie, J.A.H., Khaw, H.L., Megens, H.-J., Trinh, T.Q., Komen, H., 2020. Genotype by environment interaction between aerated and non-aerated ponds and the impact of aeration on genetic parameters in Nile tilapia (Oreochromis niloticus). Aquaculture 529, 735704. https://doi.org/10.1016/j.aquaculture.2020.735704

Meuwissen, T.H.E., Hayes, B.J., Goddard, M.E., 2001. Prediction of total genetic value using genome-wide dense marker maps. Genetics 157, 1819–1829.

Meyer, G.E., Mehta, T., Kocher, M.F., Mortensen, D.A., Samal, A., 1998. Textural imaging and discriminant analysis for distinguishingweeds for spot spraying. Transactions of the ASAE 41, 1189.

Miglior, F., Fleming, A., Malchiodi, F., Brito, L.F., Martin, P., Baes, C.F., 2017. A 100-Year Review: Identification and genetic selection of economically important traits in dairy cattle. J Dairy Sci 100, 10251–10271. https://doi.org/10.3168/jds.2017-12968

Muniesa, A., Basurco, B., Aguilera, C., Furones, D., Reverté, C., Sanjuan-Vilaplana, A., Jansen, M.D., Brun, E., Tavornpanich, S., 2020. Mapping the knowledge of the main diseases affecting sea bass and sea bream in Mediterranean. Transbound Emerg Dis 67, 1089–1100. https://doi.org/10.1111/tbed.13482

Næve, I., Korsvoll, S.A., Santi, N., Medina, M., Aunsmo, A., 2022. The power of genetics: Past and future contribution of balanced genetic selection to sustainable growth and productivity of the Norwegian Atlantic salmon (Salmo salar) industry. Aquaculture 553, 738061. https://doi.org/10.1016/j.aquaculture.2022.738061

Navarro, A., Lee-Montero, I., Santana, D., Henríquez, 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

Navarro, A., Zamorano, M.J., Hildebrandt, S., Gines, R., Aguilera, C., Afonso, J.M., 2009. Estimates of heritabilities and genetic correlations for body composition traits and G x E interactions, in gilthead seabream (Sparus auratus L.). Aquaculture 295, 183–187. https://doi.org/10.1016/j.aquaculture.2009.07.012

Nguyen, N.H., Ponzoni, R.W., Abu-Bakar, K.R., Hamzah, A., Khaw, H.L., Yee, H.Y., 2010. Correlated response in fillet weight and yield to selection for increased harvest weight in genetically improved farmed tilapia (GIFT strain), Oreochromis niloticus. Aquaculture 305, 1–5. https://doi.org/10.1016/J.AQUACULTURE.2010.04.007

Nielsen, H.M., Sonesson, A.K., Yazdi, H., Meuwissen, T.H.E., 2009. Comparison of accuracy of genome-wide and BLUP breeding value estimates in sib based aquaculture breeding schemes. Aquaculture 289, 259–264. https://doi.org/10.1016/j.aquaculture.2009.01.027

Ødegård, J., Moen, T., Santi, N., Korsvoll, S.A., Kjøglum, S., Meuwissen, T.H.E., 2014. Genomic prediction in an admixed population of Atlantic salmon (Salmo salar) . Frontiers in Genetics .

Ødegård, J., Olesen, I., Gjerde, B., Klemetsdal, G., 2006. Evaluation of statistical models for genetic analysis of challenge test data on furunculosis resistance in Atlantic salmon (Salmo salar): Prediction of field survival. Aquaculture 259, 116–123. https://doi.org/10.1016/j.aquaculture.2006.05.034

Otsu, N., 1979. A threshold selection method from gray-level histograms. IEEE Trans Syst Man Cybern 9, 62–66.

Palti, Y., Gao, G., Liu, S., Kent, M.P., Lien, S., Miller, M.R., Rexroad III, C.E., Moen, T., 2015. The development and characterization of a 57 K single nucleotide polymorphism array for rainbow trout. Mol Ecol Resour 15, 662–672.

Paolucci, M., 2023. Fish Nutrition and Feed Technology. Fishes.

Pastor, F., Valiente, J.A., Palau, J.L., 2018. Sea Surface Temperature in the Mediterranean: Trends and Spatial Patterns (1982–2016). Pure Appl Geophys 175, 4017–4029. https://doi.org/10.1007/s00024-017-1739-z

Peñaloza, C., Manousaki, T., Franch, R., Tsakogiannis, A., Sonesson, A., Aslam, M.L., Allal, F., Bargelloni, L., Houston, R.D., Tsigenopoulos, C.S., 2021. Development and testing of a combined species SNP array for the European seabass (Dicentrarchus labrax) and gilthead seabream (Sparus aurata). Genomics. https://doi.org/10.1016/j.ygeno.2021.04.038

Petrellis, N., 2021. Measurement of fish morphological features through image processing and deep learning techniques. Applied Sciences (Switzerland) 11. https://doi.org/10.3390/app11104416

Ponzoni, P.W., Nguyen, N.H., Khaw, H.L., Kamaruzzaman, N., Hamzah, A., Bakar, K.R.A., Yee, H., 2008. Genetic improvement of Nile tilapia (Oreochromis Niloticus): present and future.

Ponzoni, R.W., Nguyen, N.H., Khaw, H.L., 2007. Investment appraisal of genetic improvement programs in Nile tilapia (Oreochromis niloticus). Aquaculture 269, 187–199. https://doi.org/10.1016/J.AQUACULTURE.2007.04.054

Pook, T., Schlather, M., Simianer, H., 2020. MoBPS - Modular Breeding Program Simulator. G3 Genes|Genomes|Genetics 10, 1915–1918. https://doi.org/10.1534/g3.120.401193

Prchal, M., Bugeon, J., Vandeputte, M., Kause, A., Vergnet, A., Zhao, J., Gela, D., Genestout, L., Bestin, A., Haffray, P., Kocour, M., 2018. Potential for Genetic Improvement of the Main Slaughter Yields in Common Carp With in vivo Morphological Predictors. Front Genet.

Prchal, M., Kocour, M., Vandeputte, M., Kause, A., Vergnet, A., Zhao, J., Gela, D., Kašpar, V., Genestout, L., Bestin, A., Haffray, P., Bugeon, J., 2020. Morphological predictors of slaughter yields using 3D digitizer and their use in a common carp breeding program. Aquaculture 520, 734993. https://doi.org/10.1016/j.aquaculture.2020.734993

Quinton, C.D., Kause, A., Koskela, J., Ritola, O., 2007. Breeding salmonids for feed efficiency in current fishmeal and future plant-based diet environments. Genetics Selection Evolution 39, 431. https://doi.org/10.1186/1297-9686-39-4-431

R Core Team, 2020. R: A language and environment for statistical computing.

Robledo, D., Matika, O., Hamilton, A., Houston, R.D., 2018. Genome-Wide Association and Genomic Selection for Resistance to Amoebic Gill Disease in Atlantic Salmon. G3 Genes|Genomes|Genetics 8, 1195–1203. https://doi.org/10.1534/g3.118.200075

Rodde, C., Chatain, B., Vandeputte, M., Trinh, T.Q., Benzie, J.A.H., de Verdal, H., 2020. Can individual feed conversion ratio at commercial size be predicted from juvenile performance in individually reared Nile tilapia Oreochromis niloticus? Aquac Rep 17, 100349. https://doi.org/10.1016/j.aqrep.2020.100349

Romana-Eguia, M.R.R., Doyle, R.W., 1992. Genotype-environment interaction in the response of three strains of Nile tilapia to poor nutrition. Aquaculture 108, 1–12. https://doi.org/10.1016/0044-8486(92)90314-B

Rutten, M.J.M., Bovenhuis, H., Komen, H., 2005. Genetic parameters for fillet traits and body measurements in Nile tilapia (Oreochromis niloticus L.). Aquaculture 246, 125–132. https://doi.org/10.1016/j.aquaculture.2005.01.006

Rutten, M.J.M., Bovenhuis, H., Komen, H., 2004. Modeling fillet traits based on body measurements in three Nile tilapia strains (Oreochromis niloticus L.). Aquaculture 231, 113–122. https://doi.org/10.1016/J.AQUACULTURE.2003.11.002

Saberioon, M., Gholizadeh, A., Cisar, P., Pautsina, A., Urban, J., 2017. Application of machine vision systems in aquaculture with emphasis on fish: state-of-the-art and key issues. Rev Aquac 9, 369–387. https://doi.org/10.1111/raq.12143

Sae-Lim, P., Komen, H., Kause, A., Martin, K.E., Crooijmans, R., van Arendonk, J.A.M., Parsons, J.E., 2013. Enhancing selective breeding for growth, slaughter traits and overall survival in rainbow trout (Oncorhynchus mykiss). Aquaculture 372–375, 89–96. https://doi.org/10.1016/j.aquaculture.2012.10.031

Sae-Lim, P., Komen, H., Kause, A., Mulder, H.A., 2014. Identifying environmental variables explaining genotype-by-environment interaction for body weight of rainbow trout (Onchorynchus mykiss): reaction norm and factor analytic models. Genetics Selection Evolution 46, 16. https://doi.org/10.1186/1297-9686-46-16

Sang, N. Van, Luan, N.T., Hao, N. Van, Nhien, T. Van, Vu, N.T., Nguyen, N.H., 2020. Genotype by environment interaction for survival and harvest body weight between recirculating tank system and pond culture in Penaeus monodon. Aquaculture 525, 735278. https://doi.org/10.1016/j.aquaculture.2020.735278

Sang, N. Van, Thomassen, M., Klemetsdal, G., Gjøen, H.M., 2009. Prediction of fillet weight, fillet yield, and fillet fat for live river catfish (Pangasianodon hypophthalmus). Aquaculture 288, 166–171. https://doi.org/10.1016/J.AQUACULTURE.2008.11.030

Sasaki, Y., Miyake, T., Gaillard, C., Oguni, T., Matsumoto, M., Ito, M., Kurahara, T., Sasae, Y., Fujinaka, K., Ohtagaki, S., Dougo, T., 2006. Comparison of genetic gains per year for carcass traits among breeding programs in the Japanese Brown and the Japanese Black cattle. J Anim Sci 84, 317–323. https://doi.org/10.2527/2006.842317x

Shepherd, R.K., Kinghorn, B.P., 1993. A deterministic model of BLUP selection in two tier open nucleus breeding schemes. Livest Prod Sci 33, 341–354. https://doi.org/10.1016/0301-6226(93)90012-7

Sonesson, A.K., 2005. A combination of walk-back and optimum contribution selection in fish: a simulation study. Genetics Selection Evolution 37, 587. https://doi.org/10.1186/1297-9686-37-7-587

Sonesson, A.K., Gjerde, B., Robinson, N., 2011. A simple selection scheme to improve disease resistance and growth. Aquaculture 319, 337–341. https://doi.org/10.1016/j.aquaculture.2011.07.009

Sonesson, A.K., Meuwissen, T.H.E., 2009. Testing strategies for genomic selection in aquaculture breeding programs. Genetics Selection Evolution 41, 37. https://doi.org/10.1186/1297-9686-41-37

Sonesson, A.K., Meuwissen, T.H.E., 2000. Mating schemes for optimum contribution selection with constrained rates of inbreeding. Genetics Selection Evolution 32, 231–248. https://doi.org/10.1186/1297-9686-32-3-231

Stien, L.H., Kiessling, A., Manne, F., 2007. Rapid estimation of fat content in salmon fillets by colour image analysis. Journal of Food Composition and Analysis 20, 73–79. https://doi.org/10.1016/j.jfca.2006.07.007

Sui, J., Luan, S., Yang, G., Xia, Z., Luo, K., Tang, Q., Lu, X., Meng, X., Kong, J., 2019. Genetic parameters and selection response for the harvest body weight of the giant freshwater prawn (Macrobrachium rosenbergii) in a breeding program in China. PLoS One 14, e0218379-.

Symonds, J.E., Clarke, S.M., King, N., Walker, S.P., Blanchard, B., Sutherland, D., Roberts, R., Preece, M.A., Tate, M., Buxton, P., Dodds, K.G., 2019. Developing Successful Breeding Programs for New Zealand Aquaculture: A Perspective on Progress and Future Genomic Opportunities. Front Genet 10, 27. https://doi.org/10.3389/fgene.2019.00027

Tayamen, M.M., 2004. Nationwide dissemination of GET-EXCEL tilapia in the Philippines, in: 6th International Symposium on Tilapia in Aquaculture, ISTA. Citeseer.

Teletchea, F., Fontaine, P., 2014. Levels of domestication in fish: implications for the sustainable future of aquaculture. Fish and Fisheries 15, 181–195. https://doi.org/10.1111/faf.12006

Tine, M., Kuhl, H., Gagnaire, P.-A., Louro, B., Desmarais, E., Martins, R.S.T., Hecht, J., Knaust, F., Belkhir, K., Klages, S., Dieterich, R., Stueber, K., Piferrer, F., Guinand, B., Bierne, N., Volckaert, F.A.M., Bargelloni, L., Power, D.M., Bonhomme, F., Canario, A.V.M., Reinhardt, R., 2014. European sea bass genome and its variation provide insights into adaptation to euryhalinity and speciation. Nat Commun 5, 5770. https://doi.org/10.1038/ncomms6770

Tixier-Boichard, M., Leenstra, F., Flock, D.K., Hocking, P.M., Weigend, S., 2012. A century of poultry genetics. Worlds Poult Sci J 68, 307–321. https://doi.org/DOI: 10.1017/S0043933912000360

Trọng, T.Q., Mulder, H.A., van Arendonk, J.A.M., Komen, H., 2013. Heritability and genotype by environment interaction estimates for harvest weight, growth rate, and shape of Nile tilapia (Oreochromis niloticus) grown in river cage and VAC in Vietnam. Aquaculture 384–387, 119–127. https://doi.org/10.1016/j.aquaculture.2012.12.022

Tsai, H.-Y., Matika, O., Edwards, S.M., Antolín-Sánchez, R., Hamilton, A., Guy, D.R., Tinch, A.E., Gharbi, K., Stear, M.J., Taggart, J.B., Bron, J.E., Hickey, J.M., Houston, R.D., 2017. Genotype Imputation To Improve the Cost-Efficiency of Genomic Selection in Farmed Atlantic Salmon. G3: Genes, Genomes, Genetics 7, 1377–1383. https://doi.org/10.1534/G3.117.040717

Vallejo, R.L., Leeds, T.D., Gao, G., Parsons, J.E., Martin, K.E., Evenhuis, J.P., Fragomeni, B.O., Wiens, G.D., Palti, Y., 2017. Genomic selection models double the accuracy of predicted breeding values for bacterial cold water disease resistance compared to a traditional pedigree-based model in rainbow trout aquaculture. Genetics Selection Evolution 49, 17. https://doi.org/10.1186/s12711-017-0293-6

Van Doan, H., Hoseinifar, S.H., Sringarm, K., Jaturasitha, S., Yuangsoi, B., Dawood, M.A.O., Esteban, M.Á., Ringø, E., Faggio, C., 2019. Effects of Assam tea extract on growth, skin mucus, serum immunity and disease resistance of Nile tilapia (Oreochromis niloticus) against Streptococcus agalactiae. Fish Shellfish Immunol 93, 428–435. https://doi.org/10.1016/j.fsi.2019.07.077

van Essen, R., Mencarelli, A., van Helmond, A., Nguyen, L., Batsleer, J., Poos, J.-J., Kootstra, G., 2021. Automatic discard registration in cluttered environments using deep learning and object tracking: class imbalance, occlusion, and a comparison to human review. ICES Journal of Marine Science 78, 3834–3846.

Vandeputte, M., Bugeon, J., Bestin, A., Desgranges, A., Allamellou, J.-M., Tyran, A.-S., Allal, F., Dupont-Nivet, M., Haffray, P., 2019. First Evidence of Realized Selection Response on Fillet Yield in Rainbow Trout Oncorhynchus mykiss, Using Sib Selection or Based on Correlated Ultrasound Measurements. Front Genet.

Vandeputte, M., Puledda, A., Tyran, A.S., Bestin, A., Coulombet, C., Bajek, A., Baldit, G., Vergnet, A., Allal, F., Bugeon, J., Haffray, P., 2017. Investigation of morphological predictors of fillet and carcass yield in European sea bass (Dicentrarchus labrax) for application in selective breeding. Aquaculture 470, 40–49. https://doi.org/10.1016/j.aquaculture.2016.12.014

VanRaden, P.M., 2008. Efficient Methods to Compute Genomic Predictions. J Dairy Sci 91, 4414–4423. https://doi.org/10.3168/jds.2007-0980

Wedemeyer, G., 1996. Physiology of fish in intensive culture systems. Springer Science & Business Media.

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L.D., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., 2019. Welcome to the Tidyverse. J Open Source Softw 4, 1686.

Wright, A.I., Dunn, C.M., Hale, M., Hutchins, G.G.A., Treanor, D.E., 2020. The effect of quality control on accuracy of digital pathology image analysis. IEEE J Biomed Health Inform 25, 307–314.

Xue, Y., Bastiaansen, J.W.M., Khan, H.A., Komen, H., 2023. An analytical framework to predict slaughter traits from images in fish. Aquaculture 566, 739175. https://doi.org/https://doi.org/10.1016/j.aquaculture.2022.739175

Yáñez, J.M., Bangera, R., Lhorente, J.P., Barría, A., Oyarzún, M., Neira, R., Newman, S., 2016. Negative genetic correlation between resistance against Piscirickettsia salmonis and harvest weight in coho salmon (Oncorhynchus kisutch). Aquaculture 459, 8–13. https://doi.org/10.1016/j.aquaculture.2016.03.020

Yang, X., Zhang, S., Liu, J., Gao, Q., Dong, S., Zhou, C., 2021. Deep learning for smart fish farming: applications, opportunities and challenges. Rev Aquac 13, 66–90. https://doi.org/10.1111/raq.12464

Yoshida, G.M., de Oliveira, C.A.L., Campos, E.C., Todesco, H., Araújo, F.C.T., Karin, H.M., Zardin, A.M.S.O., Bezerra Júnior, J.S., Filho, L.A., Vargas, L., Ribeiro, R.P., 2022. A breeding program for Nile tilapia in Brazil: Results from nine generations of selection to increase the growth rate in cages. Journal of Animal Breeding and Genetics 139, 127–135. https://doi.org/10.1111/jbg.12650

Zenger, K.R., Khatkar, M.S., Jerry, D.R., Raadsma, H.W., 2017. The next wave in selective breeding: implementing genomic selection in aquaculture, in: Proc. Assoc. Advmt. Anim. Breed. Genet. pp. 105–112.

Zhou, C., Yang, X., Zhang, B., Lin, K., Xu, D., Guo, Q., Sun, C., 2017a. An adaptive image enhancement method for a recirculating aquaculture system. Sci Rep 7, 6243.

Zhou, C., Zhang, B., Lin, K., Xu, D., Chen, C., Yang, X., Sun, C., 2017b. Near-infrared imaging to quantify the feeding behavior of fish in aquaculture. Comput Electron Agric 135, 233–241.

Summary

In this PhD thesis, I focus on integrated breeding programs. In integrated breeding programs, data collection and selection are performed in the commercial production environment. The integrated approach is particularly useful for species or populations with small production volumes, where a central-nucleus breeding program is not affordable. These species deserve a breeding program as well. Selection candidates in these programs are the production animals of fish farms that are kept under specific conditions of the farm's production location. In this thesis, I aim to contribute to the improvement of integrated breeding programs for fish by addressing several challenges associated with their implementation.

In **Chapter 2**, I addressed the need to obtain phenotypes fast and accurately within integrated breeding programs. In that chapter, I validated the quality of several morphometric measurements as indicator traits for production traits, which were extracted from 2D and 3D images of gilthead seabream with a novel automated image analysis process. Automated phenotyping simplifies data collection on large numbers of fish. Phenotyping from images allows selection on indicator traits that are 1) measured on selection candidates instead of sibs and 2) have high heritabilities and correlations to breeding goal traits.

In **Chapter 3**, I addressed the challenge of controlling the costs of integrated breeding programs. In this chapter, I presented a novel genomic selection design that eliminates the need to genotype a reference population. This novel design uses images on the reference population as well as selection candidates to predict indicator traits' values for the selection candidates. Using these indicator traits as performance data reduces the cost of genotyping in comparison to traditional genomic selection designs.

In **Chapter 4**, I addressed the challenge of optimizing logistics and costs related to phenotyping. In this chapter, I investigated the balance between the selection response in harvest weight and a difficult-to-measure trait in a two-stage genomic selection design. The optimal proportion of candidates phenotyped for

a difficult-to-measure trait was investigated. The desired selection responses in harvest weight and a difficult-to-measure trait can be optimized by balancing the phenotyping effort for the difficult-to-measure trait.

In **Chapter 5**, I estimated GxE for production traits and organ weights between distinct production locations of gilthead seabream. Integrated breeding programs tailor the genetic improvement to the production system where candidates are kept. I detected strong GxE for production traits including growth and fillet yield, which is most likely because of differences in natural water conditions. This means that a balanced performance across locations may need to be recorded in integrated breeding programs if a fish farm uses multiple distinct locations to produce fish.

In **Chapter 6** (general discussion), I provide a synthesis of the knowledge produced in chapters 2-5 and how this knowledge can be used to improve integrated breeding programs for simple and affordable genetic improvement of all fish species.

About the author

Benan Gulzari was born on the 21st of January 1987 in Istanbul, Turkey. He obtained his BSc in Veterinary Science at Istanbul University in 2011 and his MSc in Aquaculture at Norwegian University of Life Sciences in 2017 with specialization in breeding and genetics. In his MSc research project, he analyzed the genetic trends in one of the oldest fish breeding programs. He is fascinated by genetics and has a passion to contribute to knowledge of genetics and genetic improvement of production populations. He pursued his passion in his PhD study at Wageningen University & Research that started in 2019. The results and implications of this PhD research project are described in this thesis.

Training and Supervision Plan (TSP)

A. The basic package	year
WIAS Introduction Day	2019
Course on philosophy of science and/or ethics	2019
Course on essential skills	2019

B. Disciplinary Competences	year
Writing a WIAS research proposal	2019
Introduction to ASReml	2019
Statistics for the life sciences	2019
Genomic predictions in animals and plants	2019
Genotype by environment interaction, uniformity and resilience	2020
Genomic predictions considering admixed populations and GxE	2022
Design of experiments	2021

C. Professional Competences	year
The essentials of scientific writing and presenting	2020
Efficient writing strategies	2020
Scientific writing	2021
The final touch	2022
Presenting with impact	2020
Critical thinking and argumentation	2020
Effective behaviour in your professional surroundings	2020
Supervising BSc and MSc thesis students	2022
Brain training	2019
Brain friendly working and writing	2020
Project and time management	2020

D. Societal Relevance	year
Societal impact of your research	2020

E. Presentation Skills	year
Oral presentation at MedAID Annual Project Meeting, Sant Carles de la Rapita, Spain	2019
Oral presentation at EAS conference, Berlin, Germany	2019
Poster presentation at WIAS Annual Conference	2020
Oral presentation at EAS conference, Online Meeting	2020
Oral presentation at WIAS Annual Conference, Online Meeting	2021
Oral presentation at WCGALP conference, Rotterdam, the Netherlands	2022

F. Teaching competences	year
Supervising an intern for MSc thesis	2020
Assisting course Genetic Improvement of Livestock	2020 - 2021

Education and Training Total (minimum 30 credits)*	**37 ECTS**

*One ECTS credit equals a study load of approximately 28 hours

Acknowledgements

It feels like yesterday when I was at Hans' office with a big smile on my face because he offered me a PhD position at Wageningen University & Research. The time that you enjoy passes very quickly. When I am writing these words, it feels like my entire PhD study has passed in an eyeblink. Although feels short, my time as a PhD candidate at WUR is filled with good memories and I can say confidently that starting a PhD study was one of the best decisions I have ever made. This does not mean that my PhD study was easy. On the contrary, it challenged me at times beyond my limits where I did not know how to tackle the problems I encountered. Although the obstacles seemed impassable, **Hans Komen** and **John Bastiaansen** illuminated the path for me that led to my personal and scientific growth. With the light they shed, I found ways to pass through the obstacles. More importantly, they equipped me with light as well so now I can illuminate the path for myself and for others. I am incredibly lucky to have them as my supervisors, whom I trust with whole my heart. They have contributed to my growth abundantly, and I cannot thank them enough for what they have given me. Their place is special in my life.

I would like to continue by acknowledging a unique person: **Kasper Janssen**. I met you when you interviewed me together with Hans for a junior researcher position within WUR. I want to thank you for welcoming me to WUR-ABG. I also want to thank you for giving me

more confidence in my work. My discussions with you were fruitful and although sometimes challenging, it was a lot of fun to work with you.

Chantal Roozeboom, I cannot finish this acknowledgement without mentioning your name. We had our weekly meetings together for almost two years. I had immense fun speaking with you in our meetings although most of them were online because of the Covid pandemic. I am grateful to you for our productive discussions and all the ideas you have given me.

Mark Camara, thank you for all your valuable contributions to my thesis. The chapters of this thesis have improved highly with your input, especially in terms of language. Primarily, you have taught me how to use English in a more concise way in scientific writing. Thanks to your efforts, my papers have become more readable and consequently, more accessible.

My special thanks to **Angelo Mencarelli**, who made things a lot easier for me by setting up an automated image analysis method. The method you created enriched my PhD study as it is the backbone of two of my chapters. I am grateful to you for contributing to an essential part of my thesis.

Thank you for truly unique data collection experiences in Greece and Spain: **Farid Aththar**, **Jan Erik Doornweerd**, **Jules Elbers**, **Koen de Reus**, **Robert Mussgnug**, **Vaibhav Mahilang**, **Varun Raj**

Nammula, **Xiaofei Yu**, and **Yangfan Liu.** Our trips to Galaxidi and El Campello were among my best memories from my PhD. I had a lot of fun working with you. We created together all the data I used in my PhD study, and I am grateful to you all for that. I also want to thank the officials of the fish farms who contributed to data collection: **Kalliopi Tsakoniti, Andromachi Gkoulia, Ernesto Prieto,** and **Marta Cortes Reyes.** My special thanks also to **Wout Abbink, Pauline Jehannet,** and **Arjan Palstra** for their efforts to train me with the ultrasound equipment.

I would like to thank **Shuwen Xia, Jani de Vos,** and **Ibrahim Jibrila** who were my neighbors at the office at different times. I enjoyed going to the office every morning knowing that I would be greeted by the nice aura you had created.

I am grateful to all members of our cozy AquaBreeding group: **Hans Komen, John Bastiaansen, Arjan Palstra, Wout Abbink, Priadi Setyawan, Farid Aththar, Pauline Jehannet, Yuuko Xue, Wisdom Agbeti, Bram Kok,** and **Meilisa Margarita.** Also, I would like to acknowledge former members: **Mark Camara, Hendrik-Jan Megens, Kasper Janssen, Samuel Mengistu, Fasil Getachew Kebede,** and **Chantal Roozeboom.**

Lisette Bourquin, Fadma Had-Dadi, Rosilde Dijkhof, and **Alex Hulzebosch**, thank you all for making life easier at the office! You have taken care of countless details related to my PhD and I am grateful to you for that.

My paranymphs **Farid Aththar** and **Yuuko Xue**, thank you for being by my side at my PhD defense!

Mariska Duenk, I want to thank you for your heartfelt creation of the cover of this thesis. The design you created is a delight to look at, and I am so satisfied that the results of 4,5 years of hard work in my PhD are presented with your art. With this design, I immortalize my gratitude to the Dutch people. The rainbow on the back of my thesis represents the welcoming and understanding nature of the Dutch people, who made me feel like I am a natural part of this nation. I am grateful to the Netherlands and the Dutch people for this. The rainbow also represents the chapters of my thesis, where each color stands for one of the chapters. When chapters are combined, they make something greater than the combination of themselves by forming a beautiful rainbow. The lighthouse in the front represents the light of research that illuminates the unseen and undiscovered of the fish. The difficulty of applying the outcome of this research is represented by the rocks in the pond. Despite difficulties, the sun starts to shine and marks the beginning of a new era!

Mother (**Safiye Süheyla Gülzari**) and father (**Rıza Gülzari**). First, thank you for granting me life so that I am here studying this PhD and pursuing my goals. Thank you for supporting me when I need your support! My sister (**Nihan Gülzari**), thank you also for supporting me when I need it!

My lovely wife Şeyda Özkan Gülzari, thank you for gifting life to our daughter and for always being by my side in good and challenging times. I am grateful to you for all the support you have given me in my life. My daughter Dünya Gülzari, you have made the world a better place with your arrival. Thank you for coming to the world and being with us!

Colophon

The research described in this thesis was financially supported by the European Commission Horizon 2020 (H2020) Framework Program through grant agreement no. 727315 for the MedAID project (Mediterranean Aquaculture Integrated Development) and no. 818367 for the AquaIMPACT.

Cover design by Mariska Duenk

Printed by Proefschriftmaken