Genomic architecture of selection for adaptation to challenging environments in Aquaculture

Xiaofei Yu

2022

Genomic architecture of selection for adaptation to challenging environments in Aquaculture

Xiaofei Yu
余小飞

Propositions

1. Genome-wide association studies are not worth the effort for animal breeding.
   (this thesis)

2. Instead of whole genome sequencing, functional annotation is required for quantitative geneticists.
   (this thesis)

3. Artificial intelligence is a challenge rather than an opportunity for biologists.

4. Understanding cultural differences is key for a successful international research career.

5. Circular agriculture will only succeed in a stable climate.

6. Farmers' needs is the first thing to consider for agricultural research.

Propositions belonging to the thesis, entitled:

Genomic architecture of selection for adaptation to challenging environments in Aquaculture

Xiaofei Yu
Wageningen, 13 December 2022

Genomic architecture of selection for adaptation to challenging environments in Aquaculture

Xiaofei Yu

Thesis committee

Promotors
Prof. Dr J. Komen
Personal Chair at Animal Breeding and Genomics
Wageningen University & Research

Prof. Dr M.A.M Groenen
Professor of Animal Breeding and Genomics
Wageningen University & Research

Co-promotor
Dr H.-J.W.C Megens
Assistant Professor, Animal Breeding and Genomics
Wageningen University & Research

Other members
Prof. Dr M. A. Peck, Wageningen University & Research
Prof. Dr P. Martinez, University of Santiago de Compostela, Spain
Dr C.S. Tsigenopoulos, Hellenic Centre for Marine Research, Greece
Dr F. Kokou, Wageningen University & Research

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

Genomic architecture of selection for adaptation to challenging environments in Aquaculture

Xiaofei Yu

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 Tuesday 13 December 2022
at 11.00 a.m. in the Omnia Auditorium.

Xiaofei Yu
Genomic architecture of selection for adaptation to challenging environments in Aquaculture
178 pages

PhD thesis, Wageningen University, Wageningen, the Netherlands (2022)
With references, with summary in English and Chinese

ISBN: 978-94-6447-469-5
DOI: https://doi.org/10.18174/579733

Abstract

Yu, X. (2022). Genomic architecture of selection for adaptation to challenging environments in Aquaculture. PhD thesis, Wageningen University and Research, the Netherlands.

Aquaculture, including freshwater and marine farming, has been important for global fish production during the past few decades. However, climate change presents a major risk threatening both quality and quantity of aquaculture production. The environmental stressors in aquaculture resulting from climate change, are temperature rise, salinity changes, sea level rise, acidification and changes of other chemical properties and changes of oxygen levels. Although a reasonable genetic gain can be achieved by selective breeding, this genetic response may not be enough to adapt fish species to the effects of climate change. Marker-assisted selection focusing on specific genes or alleles that allow fish to cope with these changes would allow more rapid adaptation of fish to these new environments. In this thesis, I focused on three essential environmental stressors - dissolved oxygen, salinity and temperature as primarily determined in aquaculture production. The main objective is to provide insight in the genomic architecture underlying the mechanism of adaptation to challenging environments of aquaculture species under farming conditions. First, I determined candidate QTL associated with phenotypic variation during adaptation to hypoxia or normoxia. I identified over-represented pathways that could explain the genetic regulation of hypoxia on growth. To identify fish with better hypoxia tolerance and growth under a hypoxic environment, I quantified the genetic correlations between an indicator trait for hypoxia tolerance (critical swimming performance) and growth. Moreover, the genomic architecture associated with swimming performance was demonstrated, while the effect of significant QTLs on growth was estimated. Beyond applying genome-wide association studies, I used selection signatures to identify QTLs and genes contributing to salinity tolerance. In addition, I also compared the genome of the saline-tolerant and highly productive tilapia “Sukamandi”, that was developed by the aquaculture research institute in Indonesia, to that of blue tilapia and Nile tilapia, to identify the QTLs contributing to salinity tolerance. Finally, I investigated QTLs associated with growth-related traits and organ weights at two distinct commercial Mediterranean product sites differing in temperature (farms in Spain and Greece). Overall, this thesis considerably adds to insight into how fish adapt to challenging environments, which will aid marker-assisted selection for improved resilience of aquaculture species under climate change.

Contents

1

General introduction

1.1 From domestication to adaptation

Domestication is defined as a process by which animals become progressively adapted to captive conditions (Price, 1997). The earliest domesticated animal was dog (*Canis familiaris*), which started at least 15,000 years ago (Savolainen et al., 2002), followed by goat, sheep, and pig between 10,500 and 10,000 B.P. (Zeder, 2006; Colledge et al., 2013; Alberto et al., 2018). Animal domestication has not only provided food, clothing and companionship, but also changed the way how people live and look at the world (Baenninger, 1995).

Compared to land animals, the domestication of fish species is much more recent (Teletchea and Fontaine, 2014). Fish domestication is initiated to control part of the life cycle, then close the entire life cycle and further establish a well-defined strain or population displaying favorable traits. Although the history of domestication is unique for every fish species, the domestication usually consists of several main steps, until populations attain the desired phenotypes as shown in **Figure 1**. The first step of fish domestication starts with transferring the fish from the wild (F0) to a new environment, usually created or controlled deliberately. Several manipulations (e.g. fishing, handling and transporting) occur during this stage, which include mechanical and physiological stressors likely to disrupt fish endocrine function. The second step of fish domestication is physiological acclimation to the captive environment, which includes a diversity of environmental factors (Lorenzen et al., 2012). Successful acclimatization requires adaptation to all these environmental factors within the tolerance range of the fish. The third step of fish domestication is to reproduce the next generation and start deliberate genetic adaptation to cultured conditions. If survival rates are sufficient, and captured fish can be successfully reproduced, the next generations can be progressively genetically adapted to the captive environment. During step 4, phenotypic changes are driven by controlled selection (selective breeding). In this phase, selection is primarily targeted at improvement of growth and other production traits.

Over the last decades, advances in genotyping and whole genome sequencing have opened up new ways of tailor-made domestication, focusing on understanding the genomic architecture of complex traits under selection. Step 5 and 6 (**Figure 1**) show how this information can be used in marker-assisted selection (MAS), which is to indirectly select for desired traits based on molecular genetic markers and Genome editing, ("CRISPR/Cas9") where genes are manipulated at the level of DNA to produce desired genotypes and phenotypes.

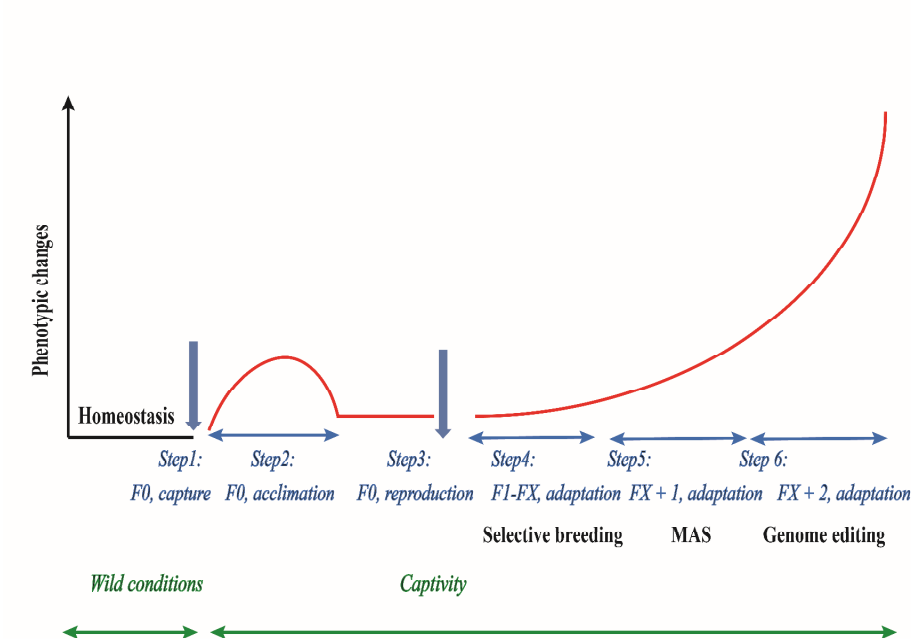


Figure 1. The main steps of fish domestication process, modified from Milla et al. (2021).

Although fish farming has been ongoing for millennia, selective breeding has only been applied recently in aquaculture. Currently only around 100 fish species are considered fully domesticated (Teletchea, 2021). The oldest domesticated fish species - common carp, dates back to about 2000 B.C. in China where carps were caught during river floods and raised to market size in artificial ponds or lagoons (Nash, 2010). Left-over fish were allowed to spawn to generate the next generation. In Europe, domestication of the common carp is thought to have started in the Danube delta from where it spread across the entire European continent during the Middle Ages (Balon, 1995). Although carp is the earliest fish species to be domesticated, it has not been subjected to selective breeding to the extent seen in the other more recently domesticated fish species (K Janssen et al., 2015).

Rainbow trout is an important, but also only recently domesticated species. Farming of Rainbow trout began in the 1870's in California. In the 1950's, several farms used artificial selection to improve traits of interest, such as body weight and early maturity, however the scientific basis of selective breeding was poorly understood at that time (Gall and Crandell, 1992). In Europe, over 20 generations of mass selection are thought to have been conducted with over 900% cumulative genetic

gain in harvest weight, which made Europe the leading Rainbow trout producer in the world (Janssen et al., 2017).

Currently the most important cultured species by value, farming of the Atlantic salmon began in the early 19th century in the United Kingdom (Ellis et al., 2016). The first selective breeding program, however, was initiated in Norway. To generate a diverse gene pool, salmon were collected from 40 Norwegian rivers and one Swedish river by AKVAFORSK in the early 1970's, to set up the first selective breeding program (Gjedrem et al., 1991). The initial breeding goal was to select candidates based on their growth performance. It has been shown that selection response for growth was about 14% per generation, and that farmed Atlantic salmon in Norway can grow twice as fast as their wild counterpart (Thodesen and Gjedrem, 2006).

Nile tilapia is considered to be one of the oldest tropical farmed species in the world, with a history dating back to more than 3500 years ago in ancient Egypt (Harache, 2002). However, selective breeding only began at the end of the 1980's with the well-documented breeding program called "Genetically Improved Farmed Tilapia" (GIFT), initiated by WorldFish with support of Norwegian and Philippine institutes. In order to have a broad genetic diversity, broodstock fish for the GIFT strain were collected from four strains of tilapia reared in the Philippines (Israel, Singapore, Taiwan, and Thailand stocks) and four wild populations from Africa (Egypt, Ghana, Kenya, and Senegal) (Gjedrem, 2012). The project aimed to select for better growth of Nile tilapia (Eknath and Acosta, 1998; Ponzoni et al., 2011).

Another popular fish species in Europe is gilthead seabream (*Sparus aurata*), which started to be cultured around 1970's in the south of Europe (Morretti, 1999). However, it was not until the early 2000's that the first commercial breeding program was initiated (Brown, 2004). To date, the reported number of selected generations varies from 1 to 5, with selection response for harvest weight of 10 – 15 % per generation can be achieved (Kasper Janssen et al., 2015).

1.2 Selective breeding and marker-assisted selection

Overall, domestication is thought to be a slow process where fish gradually adapt to the new, captive environment. By contrast, selective breeding could achieve a 7 to 25% selection response on harvest weight varied from species to species (Janssen et al., 2017). The traditional methods for selective breeding are mostly mass selection and pedigree-based selection. Mass selection is one of the earliest selection methods, where the fish are ranked based on the observed phenotype. However, mass selection does not consider family relatedness, which may result in a high

inbreeding coefficient. Selective breeding based on pedigree information is widely used as it accounts for all family relationships and fixed environmental effect simultaneously (e.g. best linear unbiased prediction: BLUP). However, keeping well-maintained pedigree records is often a problem in aquatic species. It has been reported that the selection intensity which can be achieved under family selection is actually relatively small due to problems with fertility and fecundity (Farias et al., 2017). Furthermore, reproductive biology is often limiting regarding the amount of control over mating. Since 2001, the direction in animal breeding has gradually shifted toward genomic selection, which aims to estimate a genomic breeding value by using a large number of markers across the entire genome (Meuwissen et al., 2001). So far, there are several advantages with applying genomic selection in aquaculture breeding programs (Allal and Nguyen, 2022), including 1) improvement of the genetic response and enhancing the accuracy of genomic prediction. 2) reducing the generation interval time, and 3) reducing the rate of inbreeding while increasing selection intensity.

MAS is regarded as a potentially more efficient method compared to the selective breeding methods, because it allows breeders to select individuals based on a particular genotype (Boopathi, 2020). This is of particular interest when targeting difficult traits such as health and disease resistance. Most performance traits, including growth and disease resistance, are regulated by multiple genes with small effects and few quantitative trait loci (QTL) with larger effects. Analysis of these associated genes and QTLs is essential to genetically improve fish performance. Markers can be found located within the desired genes or nearby a gene regulating the trait of interest. MAS has been shown to have several benefits : 1) select genes for traits with low heritability 2) select for recessive alleles of a particular gene, 3) understand the genomic architecture for phenotypes that are difficult, expensive and time consuming to determine 4) more precisely estimate the genetic inbreeding and manage genetic diversity (Collard and Mackill, 2008). One successful example in fish is infectious pancreatic necrosis (IPN) resistance in Atlantic salmon. IPN is a viral disease that resulted in a serious problem in salmon aquaculture especially at the beginning of the 21st century, with outbreaks causing high mortality up to 90% (Houston et al., 2020). Research teams from the UK and Norway identified a major QTL on chromosome 26 that could explain 80 to 100% of genetic variation in IPN resistance in seawater (Houston et al., 2008). The incorporation of these major QTLs in the breeding program reduced the mortality by IPN to near zero from 2009 to 2015 in Norway (Norris, 2017). Whirling disease is another pathogen-caused disease that leads to skeletal deformation and mortality of many salmonid species. A single QTL on chromosome 9 was identified which explained 50 – 86% of phenotypic

variance in rainbow trout mortality to the disease (Baerwald et al., 2011). Finally, photobacteriosis is an infectious bacterial disease and may cause high mortality (90 to 100 %) in seabream. A large genome-wide QTL was detected on LG17, which explains 13-16% of the genetic variance (Aslam et al., 2018). These examples show that the combination of selective breeding with MAS offers a great opportunity to improve selective breeding for difficult traits, such as traits which are difficult to measure and/or have low heritability. The information from these markers can be simultaneously integrated into genetic assessments to estimate the genomic breeding values.

1.3 Genomic resources for aquaculture breeding

The genomic resources for aquaculture are lagging behind those of plant and livestock, including the sequencing and assembly of references genomes (Houston et al., 2020). An important step for marker-assisted selection and genomic selection is to collect genotypes. Currently, there are three widely used methods for genotyping in aquaculture, i.e. Single Nucleotide Polymorphism (SNP) array, Genotyping-By-Sequencing (GBS), and whole genome sequencing.

SNP array is a type of DNA microarray containing probes designed to characterize polymorphisms known to exist in the species, or specifically, in a target population. The density of SNP arrays can vary from low to high depending on the experimental design. For example, the first SNP array for Nile tilapia was developed by Joshi et al. (2018), and contains more than 58,000 SNPs (58K) identified in a Genomar farmed strain. Another SNP array (65K) has been developed specifically for GIFT tilapia by the Roslin Institute. This array was subsequently validated in six other tilapia populations showing that at least 30,000 SNPs were segregating within each of these six populations (Peñaloza et al., 2020). These SNP arrays provide a good opportunity to conduct genomic selection and population management studies, e.g. assessing genetic diversity and inbreeding levels.

GBS, which targets a reduced genome representation by using restriction enzyme digestion combined with size selection, is regarded as a low-cost method for genotyping (Wickland et al., 2017). This method provides greater flexibility in terms of the targeted number of loci and aims at a genome representation from anywhere between a few tenth of a percentage to a few percentages. The molecular markers identified by GBS can be further used to address research questions in the domain of genomics and quantitative genetics. So far, GBS has been applied in a range of aquatic species summarized by Robledo et al. (2018), including *Oreochromis niloticus*, *Salmo salar*, and *Oncorhynchus mykiss*. And the number of studies based

on GBS is continuously increasing. The main advantage of GBS is that it can be applied to any species and can be done at low costs even for small experiments. Conversely, SNP arrays require extensive SNP discovery in the target species or population prior to design of the array. In addition, most of the high-density SNP array platforms only become cost effective when large batches are ordered. The downside of GBS, however, compared to SNP arrays is lower reproducibility of genotyping.

Whole genome (re-) sequencing is a comprehensive method for analyzing the entire genome, contrary to GBS which sequences a reduced genome representation. Aquaculture genome sequencing projects began in the USA, UK, some EU countries, and China in the early 2000's (Yue and Wang, 2017). Thus, the genomes of several major aquaculture species have been sequenced, including catfish, Atlantic salmon, rainbow trout, tilapia, Pacific oyster, and shrimp. There are currently several ongoing well-known genome projects related to aquaculture: the Genome 10 k Project (Koepfli et al., 2015), aiming to sequence the complete genomes of 10,000 vertebrate species, including 4000 fish species, to further understand their evolution and to rescue the endangered species; the AQUA-FAANG project (https://www.aqua-faang.eu/), aiming to improve annotation of functional genomic sequences and to use that information for genotype-to-phenotype prediction in the six most important European farmed fish species; the Fish 10 k project (Fan et al., 2020), aiming to construct reference genomes of 10,000 representative fish species and understand functioning of important issues, vertebrate evolution and fish development. Furthermore, whole genome sequence information of fish populations also allows identifying regions and genes under selection, including those involved in the domestication process (Malinsky et al., 2018). Overall, these genome resources promote not only fundamental research such as genome characterization, comparative genomics, and evolution, but also applied research in aquaculture production.

With the development of these genomic resources, QTL mapping, genome-wide association studies (GWAS), marker-assisted selection (MAS) and genomic selection (GS) are becoming routinely applied in aquaculture species. QTL mapping in aquaculture has been explored for salinity tolerance (Jiang et al., 2019), body color (Li et al., 2019), ammonia-nitrogen tolerance (Zhu et al., 2021), and sex-determination (Palaiokostas et al., 2015). QTL mapping, however, does come with limitations. For example, it requires genotyping both parents and progeny. It is also difficult to identify the specific genes that have minor effect on the target traits. By contract, a GWAS allows identifying genetic variants associated with complex traits ranging from small to large effect (see **BOX 1** for more detailed explanation).

Applying genome-wide markers of animals from a managed breeding program, it is possible to select directly for favorable alleles associated with the target traits. Also, the incorporation of whole genomic markers enables us to easily implement GS in species with advanced breeding scheme.

1.4 Climate change and aquaculture

Aquaculture, including freshwater and marine farming, now produces more seafood than wild caught fish, and this production is expected to double by 2050 (Pernet and Browman, 2021). According to the FAO (2020), aquaculture contributes a total of 82.1 million tons (46%) to global fish production, and it has continuously been rising in the past decades. However, climate change has been identified as a major risk to the global food production and a big threat to both quality and quantity of aquaculture production. The urgent concern is whether aquaculture can grow sustainably and sufficiently to meet the future demand of a rapidly growing human population given realistic climate change scenarios.

The most important threats to aquaculture resulting from global climate change are high water temperatures outside the physiological tolerance range, increases and rapid changes in salinity due to drought and flooding, acidification of seawater due to carbon dioxide uptake, and oxygen shortages caused by any combination of the above that could result in algal blooms (Myers et al., 2017; Reid et al., 2019). In this thesis, I focus on three major environmental factors: dissolved oxygen, salinity, and temperature. Because these three environmental factors primarily determine productivity in aquaculture (more explanation see each section below). My hypothesis is that, although a reasonable genetic gain can be achieved by selective breeding, this genetic response may not be enough to adapt fish species to the effects of climate change. That is because the genomic architecture of the traits underlying adaption to these environmental stressors is complex. Marker-assisted selection focusing on specific genes or alleles that confer fish adaptation could result in more rapid progress and help aquaculture cope with the challenges posed by climate change.

Fish have evolved various mechanisms, including morphological, biochemical, behavioral, and physiological, to deal with environmental stressors. Every one of these mechanisms is under genetic control, and hence phenotypic changes can be found as a consequence of genetic adaptation to extreme environments, including changes in the brain, head, and mouth of fish species (Palacios et al., 2013). Morphological and physiological adaptations are usually accompanied by additional changes such as behavior. Although non-genetic factors, often referred to as

phenotypic plasticity, can lead to adaptation to extreme environments, there is no doubt regarding the importance of genetic variation during response to different environmental stressors. Unravelling the candidate QTLs and genes underlying these novel traits, could lead to the discovery of novel markers, incorporated into breeding program for improved resilience of aquaculture species under climate change.

1.4.1 Adaptation to hypoxia stress

Dissolved oxygen (DO) is an important parameter for water quality. Hypoxia occurs when DO decreases to a harmfully low level for fish species. Fish can perform various behavioral responses to hypoxia (Chapman and Mckenzie, 2009; Domenici et al., 2013). When confronted with severe hypoxia, fish reduce movement or feed intake to conserve energy (Abdel-Tawwab et al., 2019). For species that regularly experience hypoxia stress, it is important to optimize breeding programs to select for fish that have better hypoxia tolerance. Although a simple solution seems to be that smallholder farms aerate their ponds for more oxygen supply, aeration of fish ponds is often not available or too expensive (Mengistu et al., 2020).

Fish behavior and physiological responses to hypoxia are well studied, however, the studies focused on genetics and genomics of adaptation to oxygen stress are sparse. Adaptation to oxygen stress in mammals is highly determined by genetics. Several key genes were reported for activating hypoxia-regulating pathways in mammals. For example, hypoxia-inducible factors (HIFs) are essential to regulate gene expression under hypoxia (Kietzmann et al., 2016). However, physiological adaptation to oxygen stress in fish species is far more complex and its genetics to a large extent unknown. Thus, major QTLs and genes that contribute to adaptation to hypoxia are not known.

In general, water contains less oxygen and the rate of oxygen diffusion is lower in water than in air (Dejours, 1981). In many aquatic environments, hypoxia is increasing in frequency, magnitude, and duration as a result of climate change and human activities (Verberk et al., 2022). These changes limit the oxygen uptake by fish, leaving them unable to meet their metabolic demand for oxygen. Nevertheless, a threat to fish production is unavoidable. Therefore, it is essential to select fish that survive and grow even when oxygen is limiting.

1.4.2 Adaptation to salinity

Shortages of freshwater for aquaculture have increased the need to develop aquaculture in brackish water in many countries. Rising seawater levels and prolonged periods of drought are expected to lead to salinization of groundwater,

surface water, and estuaries of coastal areas (Oppenheimer et al., 2019). Furthermore, flooding and drought could also drive large fluctuations in water salinity. Therefore, it is urgent to develop fish species for aquaculture that can adapt to brackish water, and/or fluctuating salinity levels.

Salinity is specific to the aquaculture environment. Extensive studies have been conducted on fish physiological responses under different levels of salinity. Euryhaline fish – fish that can deal with a wide range of salinities - keep a stable internal osmotic balance through the osmoregulatory systems involving gills, kidneys, and intestine (Edwards and Marshall, 2012). When the salinity exceeds the tolerance of the fish, inflammation, apoptosis and other negative consequences can occur (Khairnar et al., 2015). Previous studies reported that genetics may play a key role in adaptation to environmental salinity challenges. For example, genes coding for prolactin (Breves et al., 2013), growth hormone (Deane and Woo, 2009), and insulin-like growth factor 1 (Yan et al., 2020) were found to contribute to salinity tolerance.

The salinity tolerance of fish varies among species and hybrids. The most widely culture tilapia – Nile tilapia (*Oreochromis niloticus*), is only moderately saline-tolerant. A significant growth decrease was observed for this species when salinity is higher than 8 ppt (parts per thousand). Other species in the genus, notably *Oreochromis* aureus and *Oreochromis mossambicus,* are saline-tolerant tilapias (Kamal and Mair, 2005; Gan et al., 2016). Parental *O. mossambicus* has a higher salinity tolerance and *O. niloticus* has a lower salinity tolerance compared to their F1 hybrids, and the salinity tolerance can increase in offspring if they are backcrossed with the parental *O. mossambicus* (Baroiller and Rogñon, 2004). Such crosses can also show significant heterosis, which could potentially lead to better growth performance than parental *O. mossambicus*. Thus, a saline-tolerant and highly productive tilapia can be produced by selective breeding of species hybrids. In this thesis we use the Sukamandi strain developed by Research Institute for Fish Breeding (RIFB) in Indonesia. This strain of hybrid origin shows increased salinity tolerance but how the underlying genomic architecture was shaped by artificial selection for a better adaptation to salinity tolerance is unclear.

1.4.3 Adaption to temperature

Global temperature obviously has been increasing and continues to increase even more due to climate change (Iona et al., 2018). Aquaculture losses and extreme temperature events are linked. Some regions with the highest concern are the largest aquaculture producing nations. For example, China, Vietnam, and part of the

Mediterranean (i.e. Spain), are predicted to be highly influenced by extreme climatic events (Reid et al., 2019). During marine heatwaves ocean temperatures rise over five degrees centigrade above normal. This can lead to mass mortalities over a period of a few days, as was seen in New Zeeland where high water temperatures caused close to 1300 tons of dead salmon during the summer of 2021 based on the SPINOFF report (https://thespinoff.co.nz/). During summer, heatwaves in the Mediterranean have led to massive fish mortalities in aquaculture facilities and the wild (Rosa et al., 2012).

The general principles of fish thermal physiology and stress responses have been extensively studied (Faught et al., 2019; Alfonso et al., 2021). For most species, body temperatures equal water temperatures and thus there is limited opportunity for self-regulation independent of water temperature. Changes in water temperature can have pronounced effect on fish physiology, such as muscle and cardiovascular function, movement, and growth (Little et al., 2020). Like hypoxia and salinity challenges, the response to a temperature challenge is also mediated by genetics. Several genes were reported to play an important role in adaptation during temperature challenge, such as Nuclear Protein-1 (*Nupr1*) and parkin E3 ubiquitin protein ligase (*Park2*), to the extent that they are considered thermal stress markers (Hori et al., 2010).

1.5 Objectives and outline of this thesis

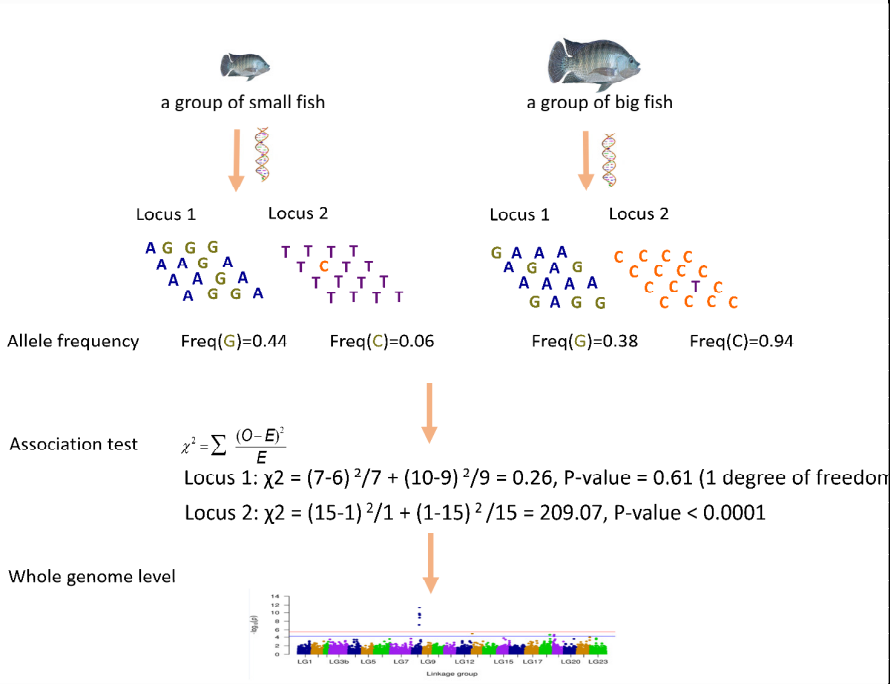
Over the past decade, an increasing number of GWAS have been conducted for aquaculture species, however mostly focusing on the traits related to growth, disease resistance and sexual maturation (Palaiokostas and Houston, 2017). The genomic architectures underlying adaptation to challenging environments, i.e., low oxygen, fluctuating or high salinity and increasing temperatures, remain unclear.

In this thesis, I aim to provide insight into the genomic architecture underlying the mechanism of adaptation to challenging environments of aquaculture species under farming conditions. I specifically focus on two fish species which experience environmental variability in three essential environmental factors: dissolved oxygen, salinity, and temperature. The ultimate objective is to provide the genomic basis of marker-assisted selection that should aid in breeding more resilient and robust fish under challenging environments.

Box 1 The principle of genome-wide association study

The genetic variation is the major cause for difference between individuals. A genome-wide association study (GWAS) is a statistical method to identify the associations between genetic variation and phenotype. Single nucleotide polymorphisms (SNPs) are the most common type of genetic variation. And a GWAS is used to detect the association between genotype and phenotype.

Example: two groups of 16 small and 16 big Nile tilapia are sampled. Genotype information is captured through a genotyping array. Thus, allele frequency of the G allele for locus 1 is estimated as 0.44 (7/16) and 0.38 (6/16) for small and big fish populations, respectively. A chi-square (χ^2) test shows that χ^2 are 0.26 and 209.07 for locus 1 and 2, respectively. The probability 0.61 suggest that there is no association between locus 1 and fish size, while the probability less than 0.0001 suggest that locus 2 is associated with fish size. Therefore, Locus 2 is a candidate marker for body size (fish photo adopted from https://stock.adobe.com/).



Noteworthily, the significance at the whole genome level needs to be determined by multiple testing corrections to avoid false positives. Furthermore, population structure needs to be accounted for performing GWAS with many individuals.

A graphical outline of the thesis and the individual objectives is shown in **Figure 2**. The experimental basis for **Chapters 2** and **3** is Nile tilapia (the GIFT strain) from WorldFish at the Aquaculture Extension Centre in Jitra (Kedah, Malaysia). In **Chapter 2**, the objective is to understand the genomic architecture associated with phenotypic variation during adaptation to hypoxia or normoxia, and further elucidate the effect of hypoxia on the genetic regulation of growth. I apply a genome-wide association study (GWAS) to identify genetic variants associated with growth traits under normoxic and hypoxic environments, while the important genes and over-represented pathways for better growth under hypoxic environment are unraveled. In **Chapter 3**, the objective is to investigate the genomic architecture of a potential indicator trait, critical swimming performance or U_{crit}, for hypoxia tolerance, and quantify the correlations with the growth under hypoxic environment. U_{crit} is a novel trait that determines the moment when fish become fatigued and stop swimming during a swimming challenge at increasing flow rates. U_{crit} may reflect the oxygen uptake efficiency and cardio-respiratory health and therefore the potential for growth under challenging oxygen conditions. I estimate the heritability of swimming performance and the genetic correlations between U_{crit} with growth traits at an early life stage, and with growth traits at a later life stage using a genomic relationship matrix under hypoxic environment. Furthermore, I identify QTLs associated with U_{crit} and estimate the significant QTLs effect on growth.

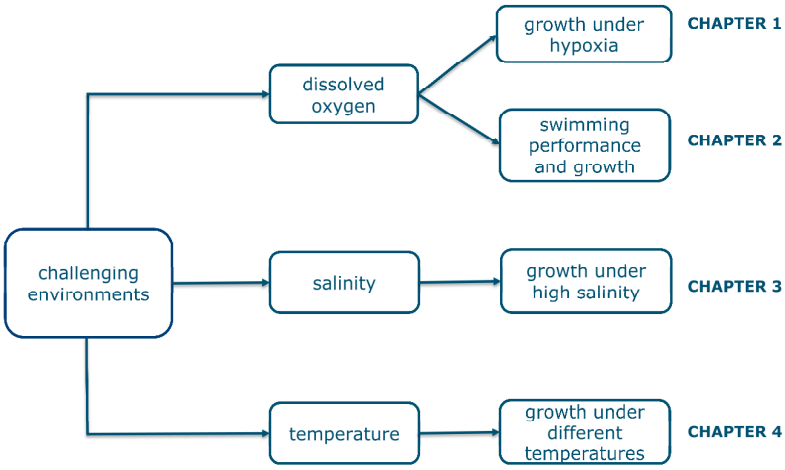


Figure 2. The outline of this thesis with specific objectives of the individual chapter.

A saline-tolerant tilapia was developed by the aquaculture research institute, Research Institute for Fish Breeding (RIFB) in Indonesia. This strain, called

"Sukamandi" shows a stable growth under high and fluctuating salinity varying from 30 to 58 ppt. The history of this strain, however, is not fully documented, but it is assumed to be derived from a hybrid between Nile tilapia and blue tilapia. In **Chapter 4**, my objective is to identify the candidate QTLs and genes associated with salinity tolerance. Instead of applying GWAS, I determined potential genes contributing to salinity tolerance from signatures of selection with whole genome sequencing data. Since the strain has been selected over four generations under a saline environment, genome contributions of blue tilapia (*Oreochromis aureus*) and Nile tilapia (*Oreochromis niloticus*) can be compared to identify the genomic regions contributing to salinity tolerance. In **Chapter 5**, the objective is to identify the genomic architecture associated with growth-related traits and organ weights under two different temperature environments. Gilthead seabream is an important fish species for Mediterranean aquaculture. More than 80% of the genetically improved fingerlings of this species originate from a single country (Greece). With the observed temperature divergence in the Mediterranean region, I investigate the QTLs associated with growth-related traits and organ weights growing out in two distinct commercial product sites (Spain and Greece farms), and further identify the over-represented processes and pathways underlying temperature specific environment. Finally, in **Chapter 6**, the overall findings relevant to this thesis are synthesized. I also propose the future directions and give concluding remarks of the whole thesis.

References

Abdel-Tawwab, M., Monier, M.N., Hoseinifar, S.H., Faggio, C., 2019. Fish response to hypoxia stress: growth, physiological, and immunological biomarkers. Fish physiology and biochemistry 45, 997–1013.

Alberto, F.J., Boyer, F., Orozco-terWengel, P., Streeter, I., Servin, B., de Villemereuil, P., Benjelloun, B., Librado, P., Biscarini, F., Colli, L., 2018. Convergent genomic signatures of domestication in sheep and goats. Nature communications 9, 1–9.

Alfonso, S., Gesto, M., Sadoul, B., 2021. Temperature increase and its effects on fish stress physiology in the context of global warming. Journal of Fish Biology 98, 1496–1508.

Allal, F., Nguyen, N.H., 2022. Genomic Selection in AquacultureAquaculture Species BT - Complex Trait Prediction: Methods and Protocols, in: Ahmadi, N., Bartholomé, J. (Eds.), . Springer US, New York, NY, pp. 469–491. https://doi.org/10.1007/978-1-0716-2205-6_17

Aslam, M.L., Carraro, R., Bestin, A., Cariou, S., Sonesson, A.K., Bruant, J.S., Haffray, P., Bargelloni, L., Meuwissen, T.H.E., 2018. Genetics of resistance to photobacteriosis in gilthead sea bream (*Sparus aurata*) using 2b-RAD sequencing. BMC Genetics 19, 1–14. https://doi.org/10.1186/s12863-018-0631-x

Baenninger, R., 1995. Some Consequences of Animal Domestication for Humans. Anthrozoös 8, 69–77. https://doi.org/10.2752/089279395787156383

Baerwald, M.R., Petersen, J.L., Hedrick, R.P., Schisler, G.J., May, B., 2011. A major effect quantitative trait locus for whirling disease resistance identified in rainbow trout (*Oncorhynchus mykiss*). Heredity 106, 920–926. https://doi.org/10.1038/hdy.2010.137

Balon, E.K., 1995. Origin and domestication of the wild carp, *Cyprinus carpio*: from Roman gourmets to the swimming flowers. Aquaculture 129, 3–48.

Baroiller, J.F., Rogñon, X., 2004. Salinity tolerance of *Oreochromis niloticus* and *O. mossambicus* F1 hybrids and their successive backcross 426–438.

Boopathi, N.M., 2020. Genetic mapping and marker-assisted selection: setting the background, in: genetic mapping and marker assisted selection. Springer, pp. 1–20.

Breves, J.P., Serizier, S.B., Goffin, V., McCormick, S.D., Karlstrom, R.O., 2013. Prolactin regulates transcription of the ion uptake Na+/Cl− cotransporter (ncc) gene in zebrafish gill. Molecular and cellular endocrinology 369, 98–106.

Brown, R.C., 2004. Genetic management and selective breeding in farmed populations of gilthead seabream (*Sparus aurata*).

Chapman, L.J., Mckenzie, D.J., 2009. Behavioral responses and ecological consequences, in: Fish Physiology. Elsevier, pp. 25–77.

Collard, B.C.Y., Mackill, D.J., 2008. Marker-assisted selection: an approach for precision plant breeding in the twenty-first century. Philosophical Transactions of the Royal Society B: Biological Sciences 363, 557–572.

Colledge, S., Conolly, J., Dobney, K., Manning, K., Shennan, S., 2013. Origins and spread of domestic animals in southwest Asia and Europe. Left Coast Press.

Deane, E.E., Woo, N.Y.S., 2009. Modulation of fish growth hormone levels by salinity, temperature, pollutants and aquaculture related stress: A review. Reviews in Fish Biology and Fisheries 19, 97–120. https://doi.org/10.1007/s11160-008-9091-0

Dejours, P., 1981. Principles of comparative respiratory physiology. sole distributors for the USA and Canada, Elsevier North-Holland.

Domenici, P., Herbert, N.A., Lefrançois, C., Steffensen, J.F., McKenzie, D.J., 2013. The effect of hypoxia on fish swimming performance and behaviour, in: Swimming Physiology of Fish. Springer, pp. 129–159.

Edwards, S.L., Marshall, W.S., 2012. Principles and patterns of osmoregulation and

euryhalinity in fishes, in: Fish Physiology. Elsevier, pp. 1–44.
Eknath, A.E., Acosta, B.O., 1998. Genetic improvement of farmed tilapias (GIFT) project: Final report, March 1988 to December 1997.
Ellis, T., Turnbull, J.F., Knowles, T.G., Lines, J.A., Auchterlonie, N.A., 2016. Trends during development of Scottish salmon farming: An example of sustainable intensification? Aquaculture 458, 82–99.
Fan, G., Song, Y., Yang, L., Huang, X., Zhang, S., Zhang, M., Yang, X., Chang, Y., Zhang, H., Li, Y., 2020. Initial data release and announcement of the 10,000 Fish Genomes Project (Fish10K). GigaScience 9, giaa080.
FAO, 2020. state of World fisheries and aquaculture 2020 : sustainability in action LK - https://wur.on.worldcat.org/oclc/1312812320.
Farias, T., César, J., Silva, L., 2017. Methods of selection using the quantitative genetics in aquaculture-a short review. Insights Aquacult. Biotech 1, 1–8.
Faught, E., Hernandez-Perez, J., Wilson, J.M., Vijayan, M.M., 2019. Stress in response to environmental changes. Climate change and non-infectious fish disorders.
Gall, G. A. E., Crandell P.A., 1992. The rainbow trout. Aquaculture 100, 1–10.
Gan, L., Xu, Z.X., Ma, J.J., Xu, C., Wang, X.D., Chen, K., Chen, L.Q., Li, E.C., 2016. Effects of salinity on growth, body composition, muscle fatty acid composition, and antioxidant status of juvenile Nile tilapia *Oreochromis niloticus* (Linnaeus, 1758). Journal of Applied Ichthyology 32, 372–374.
Gjedrem, T., 2012. Genetic improvement for the development of efficient global aquaculture: a personal opinion review. Aquaculture 344, 12–22.
Gjedrem, T., Gjøen, H.M., Gjerde, B., 1991. Genetic origin of Norwegian farmed Atlantic salmon. Aquaculture 98, 41–50. https://doi.org/10.1016/0044-8486(91)90369-I
Harache, Y., 2002. Development and diversification issues in aquaculture. A historical and dynamic view of fish culture diversification. Cahiers Options Méditerranéennes 59, 15–23.
Hori, T.S., Gamperl, A.K., Afonso, L.O.B., Johnson, S.C., Hubert, S., Kimball, J., Bowman, S., Rise, M.L., 2010. Heat-shock responsive genes identified and validated in Atlantic cod (*Gadus morhua*) liver, head kidney and skeletal muscle using genomic techniques. BMC genomics 11, 1–22.
Houston, R.D., Bean, T.P., Macqueen, D.J., Gundappa, M.K., Jin, Y.H., Jenkins, T.L., Selly, S.L.C., Martin, S.A.M., Stevens, J.R., Santos, E.M., 2020. Harnessing genomics to fast-track genetic improvement in aquaculture. Nature Reviews Genetics 21, 389–409.
Houston, R.D., Haley, C.S., Hamilton, A., Guy, D.R., Tinch, A.E., Taggart, J.B., McAndrew, B.J., Bishop, S.C., 2008. Major quantitative trait loci affect resistance to infectious pancreatic necrosis in Atlantic salmon (*Salmo salar*). Genetics 178, 1109–1115.
Iona, A., Theodorou, A., Sofianos, S., Watelet, S., Troupin, C., Beckers, J.-M., 2018. Mediterranean Sea climatic indices: monitoring long-term variability and climate changes. Earth System Science Data 10, 1829–1842.
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, Kasper, 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. Citeseer.
Janssen, K, Prchal, M., Kocour, M., Berentsen, P.B.M., Komen, H., 2015. Common Carp—Current Status of Selective Breeding in Europe.
Jiang, D.L., Gu, X.H., Li, B.J., Zhu, Z.X., Qin, H., Meng, Z. ning, Lin, H.R., Xia, J.H., 2019. Identifying

a Long QTL Cluster Across chrLG18 Associated with Salt Tolerance in Tilapia Using GWAS and QTL-seq. Marine Biotechnology 21, 250–261. https://doi.org/10.1007/s10126-019-09877-y

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*). Frontiers in Genetics 9. https://doi.org/10.3389/fgene.2018.00472

Kamal, A.H.M.M., Mair, G.C., 2005. Salinity tolerance in superior genotypes of tilapia, *Oreochromis niloticus*, *Oreochromis mossambicus* and their hybrids, in: Aquaculture. pp. 189–201. https://doi.org/10.1016/j.aquaculture.2005.02.008

Khairnar, S.O., Tian, X., Fang, Z., Dong, S., 2015. Effects of the amplitude and frequency of salinity fluctuation on the body composition and energy budget of juvenile tongue sole (*Cynoglossus semilaevis*). Journal of Ocean University of China 14, 127–134.

Kietzmann, T., Mennerich, D., Dimova, E.Y., 2016. Hypoxia-inducible factors (HIFs) and phosphorylation: impact on stability, localization, and transactivity. Frontiers in cell and developmental biology 4, 11.

Koepfli, K.-P., Paten, B., Scientists, G. 10K C. of, O'Brien, S.J., 2015. The Genome 10K Project: a way forward. Annu. Rev. Anim. Biosci. 3, 57–111.

Li, B.J., Zhu, Z.X., Gu, X.H., Lin, H.R., Xia, J.H., 2019. QTL Mapping for Red Blotches in Malaysia Red Tilapia (*Oreochromis spp.*). Marine Biotechnology 384–395. https://doi.org/10.1007/s10126-019-09888-9

Little, A.G., Loughland, I., Seebacher, F., 2020. What do warming waters mean for fish physiology and fisheries? Journal of Fish Biology 97, 328–340.

Lorenzen, K., Beveridge, M.C.M., Mangel, M., 2012. Cultured fish: integrative biology and management of domestication and interactions with wild fish. Biological Reviews 87, 639–660.

Malinsky, M., Svardal, H., Tyers, A.M., Miska, E.A., Genner, M.J., Turner, G.F., Durbin, R., 2018. Whole-genome sequences of Malawi cichlids reveal multiple radiations interconnected by gene flow. Nature Ecology and Evolution 2, 1940–1955. https://doi.org/10.1038/s41559-018-0717-x

Mengistu, S.B., Mulder, H.A., Benzie, J.A.H., Komen, H., 2020. A systematic literature review of the major factors causing yield gap by affecting growth, feed conversion ratio and survival in Nile tilapia (Oreochromis niloticus). Reviews in Aquaculture 12, 524–541. https://doi.org/10.1111/raq.12331

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

Milla, S., Pasquet, A., El Mohajer, L., Fontaine, P., 2021. How domestication alters fish phenotypes. Reviews in Aquaculture 13, 388–405.

Morretti, A., 1999. Manual on hatchery production of seabass and gilthead seabream. Food & Agriculture Org.

Myers, S.S., Smith, M.R., Guth, S., Golden, C.D., Vaitla, B., Mueller, N.D., Dangour, A.D., Huybers, P., 2017. Climate Change and Global Food Systems: Potential Impacts on Food Security and Undernutrition. Annual Review of Public Health 38, 259–277. https://doi.org/10.1146/annurev-publhealth-031816-044356

Nash, C., 2010. The history of aquaculture. John Wiley & Sons.

Norris, A., 2017. Application of genomics in salmon aquaculture breeding programs by Ashie Norris: Who knows where the genomic revolution will lead us? Marine Genomics 36, 13–15. https://doi.org/10.1016/j.margen.2017.11.013

Oppenheimer, M., Glavovic, B., Hinkel, J., van de Wal, R., Magnan, A.K., Abd-Elgawad, A., Cai, R., Cifuentes-Jara, M., Deconto, R.M., Ghosh, T., 2019. Sea level rise and implications for low lying islands, coasts and communities.

Palacios, M., Arias-Rodriguez, L., Plath, M., Eifert, C., Lerp, H., Lamboj, A., Voelker, G., Tobler, M., 2013. The Rediscovery of a Long Described Species Reveals Additional Complexity in Speciation Patterns of Poeciliid Fishes in Sulfide Springs. PLoS ONE 8. https://doi.org/10.1371/journal.pone.0071069

Palaiokostas, C., Bekaert, M., Khan, M.G.Q., Taggart, J.B., Gharbi, K., McAndrew, B.J., Penman, D.J., 2015. A novel sex-determining QTL in Nile tilapia (*Oreochromis niloticus*). BMC Genomics 16, 1–10. https://doi.org/10.1186/s12864-015-1383-x

Palaiokostas, C., Houston, R.D., 2017. Genome-wide approaches to understanding and improving complex traits in aquaculture species. CAB Reviews: Perspectives in Agriculture, Veterinary Science, Nutrition and Natural Resources 12, 1–10. https://doi.org/10.1079/PAVSNNR201712055

Peñaloza, C., Robledo, D., Barría, A., Trinh, T.Q., Mahmuddin, M., Wiener, P., Benzie, J.A.H., Houston, R.D., 2020. Development and validation of an open access SNP array for Nile Tilapia (*Oreochromis niloticus*). G3: Genes, Genomes, Genetics 10, 2777–2785. https://doi.org/10.1534/g3.120.401343

Pernet, F., Browman, H.I., 2021. The future is now: marine aquaculture in the anthropocene. ICES Journal of Marine Science.

Ponzoni, R.W., Nguyen, N.H., Khaw, H.L., Hamzah, A., Bakar, K.R.A., Yee, H.Y., 2011. Genetic improvement of Nile tilapia (*Oreochromis niloticus*) with special reference to the work conducted by the World Fish Center with the GIFT strain. Reviews in Aquaculture 3, 27–41. https://doi.org/10.1111/j.1753-5131.2010.01041.x

Price, E., 1997. Behavioural genetics and the process of animal domestication. Genetics and the behaviour of domestic animals 31–65.

Reid, G.K., Gurney-Smith, H.J., Marcogliese, D.J., Knowler, D., Benfey, T., Garber, A.F., Forster, I., Chopin, T., Brewer-Dalton, K., Moccia, R.D., 2019. Climate change and aquaculture: considering biological response and resources. Aquaculture Environment Interactions 11, 569–602.

Robledo, D., Palaiokostas, C., Bargelloni, L., Martínez, P., Houston, R., 2018. Applications of genotyping by sequencing in aquaculture breeding and genetics. Reviews in Aquaculture 10, 670–682. https://doi.org/10.1111/raq.12193

Rosa, R., Marques, A., Nunes, M.L., 2012. Impact of climate change in Mediterranean aquaculture. Reviews in Aquaculture 4, 163–177.

Savolainen, P., Zhang, Y., Luo, J., Lundeberg, J., Leitner, T., 2002. Genetic evidence for an East Asian origin of domestic dogs. Science 298, 1610–1613.

Teletchea, F., 2021. Fish domestication in aquaculture: 10 unanswered questions. Animal Frontiers 11, 87–91.

Teletchea, F., Fontaine, P., 2014. Levels of domestication in fish: implications for the sustainable future of aquaculture. Fish and fisheries 15, 181–195.

Thodesen, J., Gjedrem, T., 2006. Breeding programs on Atlantic salmon in Norway: lessons learned.

Verberk, W.C.E.P., Sandker, J.F., van de Pol, I.L.E., Urbina, M.A., Wilson, R.W., McKenzie, D.J., Leiva, F.P., 2022. Body mass and cell size shape the tolerance of fishes to low oxygen in a temperature-dependent manner. Global Change Biology.

Wickland, D.P., Battu, G., Hudson, K.A., Diers, B.W., Hudson, M.E., 2017. A comparison of genotyping-by-sequencing analysis methods on low-coverage crop datasets shows advantages of a new workflow, GB-eaSy. BMC bioinformatics 18, 1–12.

Yan, J.J., Lee, Y.C., Tsou, Y.L., Tseng, Y.C., Hwang, P.P., 2020. Insulin-like growth factor 1 triggers salt secretion machinery in fish under acute salinity stress. Journal of Endocrinology 246, 277–288. https://doi.org/10.1530/JOE-20-0053

Yue, G.H., Wang, L., 2017. Current status of genome sequencing and its applications in

aquaculture. Aquaculture 468, 337–347. https://doi.org/10.1016/j.aquaculture.2016.10.036

Zeder, M.A., 2006. A critical assessment of markers of initial domestication in goats (*Capra hircus*). Documenting domestication: New genetic and archaeological paradigms 181–208.

Zhu, Z.X., Lin, Y.L., Qin, H., Xiong, Y.Y., Jiang, D.L., Lin, H.R., Yu, Z.L., Xia, J.H., 2021. Identifying a genome-wide QTL interval controlling for ammonia-nitrogen tolerance on chrLG1 of Nile tilapia. Aquaculture 543, 736946. https://doi.org/10.1016/j.aquaculture.2021.736946

2

Genome-wide association analysis of adaptation to oxygen stress in Nile tilapia (*Oreochromis niloticus*)

Xiaofei Yu[1], Hendrik-Jan Megens[1], Samuel Bekele Mengistu[1,2], John W.M. Bastiaansen[1], Han A. Mulder[1], John A.H. Benzie[3,4], Martien A.M. Groenen[1], Hans Komen[1]

[1] Animal Breeding and Genomics, Wageningen University & Research, The Netherlands
[2]School of Animal and Range Sciences, College of Agriculture, Hawassa University, Hawassa Ethiopia
[3]WorldFish Centre, Jalan Batu Maung, Bayan Lepas, Penang, Malaysia
[4]School of Biological Earth and Environmental Sciences, University College Cork, Cork, Ireland

BMC genomics 22.1 (2021): 1-13.

Abstract

Background: Tilapia is one of the most abundant species in aquaculture. Hypoxia is known to depress growth rate, but the genetic mechanism by which this occurs is unknown. In this study, two groups consisting of 3140 fish were raised in either aerated (normoxia) or non-aerated pond (nocturnal hypoxia). During grow out, fish were sampled five times to determine individual body weight (BW) gains. We applied a genome-wide association study to identify SNPs and genes associated with the hypoxic and normoxic environments in the 16th generation of a Genetically Improved Farmed Tilapia population.

Results: In the hypoxic environment, 36 SNPs associated with at least one of the five body weight measurements (BW1 till BW5), of which six, located between 19.48 Mb and 21.04 Mb on Linkage group (LG) 8, were significant for body weight in the early growth stage (BW1 to BW2). Further significant associations were found for BW in the later growth stage (BW3 to BW5), located on LG1 and LG8. Analysis of genes within the candidate genomic region suggested that MAPK and VEGF signalling were significantly involved in the later growth stage under the hypoxic environment. Well-known hypoxia-regulated genes such as *igf1rb*, *rora*, *efna3* and *aurk* were also associated with growth in the later stage in the hypoxic environment. Conversely, 13 linkage groups containing 29 unique significant and suggestive SNPs were found across the whole growth period under the normoxic environment. A meta-analysis showed that 33 SNPs were significantly associated with BW across the two environments, indicating a shared effect independent of hypoxic or normoxic environment. Functional pathways were involved in nervous system development and organ growth in the early stage, and oocyte maturation in the later stage.

Conclusions: There are clear genotype-growth associations in both normoxic and hypoxic environments, although genome architecture involved changed over the growing period, indicating a transition in metabolism along the way. The involvement of pathways important in hypoxia especially at the later growth stage indicates a genotype-by-environment interaction, in which MAPK and VEGF signalling are important components.

Keywords: Nile tilapia, growth, hypoxia, oxygen stress, GWAS, meta-analysis

2.1 Introduction

Tilapia is one of the most important species in aquaculture noted for their relative ease of culture and rapid growth. Tilapia is currently cultured in over 120 countries, mainly in the tropics and sub-tropics, with a production from 0.3 million tonnes in 1987 to closely 7 million tonnes in 2018, which makes it the second largest aquaculture species in the world [1]. Tilapia is a valuable protein source in developing and emerging economies. Due to its wide range of culturing conditions, tilapia is also an excellent model to study adaptive responses to environmental stresses [2]. One of the most important non-commercial breeding programs is the Genetically Improved Farmed Tilapia (GIFT), executed by WorldFish in Malaysia. It has sustained genetic gains for growth and body trait more than 10% per generation for more than six generations [3]. However, rapid growth potentially exacerbates existing limitations in the production environment. In non-aerated ponds, high stocking density can lead to an extreme hypoxic environment, especially at the end of the night (nocturnal hypoxia), when algae have higher rate of oxygen consumption than oxygen production. The extreme hypoxic environment can lead to lower feed intake, stagnated growth, and susceptibility to disease [4, 5]. The result is a higher mortality and lower yield than what could potentially be achieved [6]. The effects can be mitigated through mechanical aeration of ponds, but a daily fluctuation in oxygen availability is nevertheless inevitable.

Response to hypoxia is a highly complicated biological process that has received considerable scientific attention, both in fishes and in land vertebrates (e.g. high-altitude adaptation studies). Most of these response processes happen very early at the onset of hypoxia through the activation of pathways depending on proteins that are already present [7]. But in the longer term, adaptive responses to hypoxia are leading to different expression of genes. In mammals, studies in the past decades pointed to an essential role of the hypoxia-inducible factors (HIF) for gene expression regulation during hypoxia [8]. Other genes such as tyrosine hydroxylase (TH), phosphoglycerate kinase 1 (PGK1) and vascular endothelial growth factor (VEGF) are also important key actors [9]. Recent studies have described that fish have homologs of HIF-α and -β, which may show similar function to those in mammals in the hypoxic environment [9, 10]. Several other hypoxia-related proteins and signal pathways have been reported, such as AMP-activated protein kinase (AMPK), reactive oxygen species (ROS), mitogen-activated protein kinase (MAPK) and IGF-1/PI3K/AKT signalling, which have been reported to be involved in hypoxia adaptation of some fish species [11, 12].

Genetic adaptation to hypoxia is important for survival in many aquatic species, since variation in oxygen availability in water can vary far more, and far more rapidly, than in terrestrial ecosystems. Hypoxia is an important cause of economic losses in aquaculture. Understanding the genomic architecture of hypoxia adaptation could help to improve resilience through breeding programs for economically important species. So far, hypoxia tolerance has been studied in a limited number of fish species, including catfish [13, 14], Atlantic salmon [15], and tilapia [16], with the aim to identify QTLs for hypoxia-tolerant traits. Genome-wide association study (GWAS) has been regarded as a powerful tool to identify genetic markers associated with target traits, and a more complete gene network will provide the knowledge bases required for the aquaculture industry to make improvements [17]. In hybrid catfish, Zhong et al. [13] revealed in total nine SNPs associated with dissolved oxygen (DO) level using a 250K SNP array. Analysis of the genes overlapping or close to those SNPs suggested that many of those genes were involved in the PI3K/AKT and VEGF pathways. In another study, Brennan et al. [18] aimed to identify population differences in hypoxia tolerance by calculating the amount of time for Killifish to lose equilibrium using GWAS. They found that variation in Hyaluronan synthase 1 (*has1*) influenced the production of hyaluronan, which can directly affect hypoxia tolerance.

There are only a few studies that focused on genetic bases of either hypoxia tolerance or growth in Nile tilapia [16, 19], however, none of these investigated how hypoxia influences growth in Nile tilapia. The main objective of this study was to unravel the genomic architecture associated with phenotypic variation during adaptation to hypoxia or normoxia, and to elucidate the effect of hypoxia on the genetic regulation of growth.

2.2 RESULTS

2.2.1 Phenotype Statistics

Fish fry was produced from generation 15 of the GIFT breeding program. The experiment was carried out in an aerated (normoxic) and non-aerated (nocturnal hypoxic) ponds, each producing 1026 and 1037 fish that were involved in the analysis. Body weight of growing fish was measured at five time points (Table 1). The data show that the number of tilapia in both environments gradually decreased. This effect was more pronounced in the hypoxic environment, with a total loss from stocking to harvest of 23% of the initial number of individuals, compared to 14% in the normoxic environment. The average body weight at five time points in the normoxic environment was significantly higher than those in the hypoxic environment, with the exception of the first time point (BW1). Interestingly, the

coefficient of variation in body weight (CV) at each time point in the two separate environments decreased.

The estimated phenotypic correlations for body weight between different time points in the two environments are shown in Table 2. Results show that phenotypic correlation between time points in the hypoxic and nomorxic environments was initially high (0.80 and 0.81 separately), but decreased with increasing time between measurements.

Table 1 Summary statistics of body weight across the whole growth period in Nile tilapia

Trait	Days	Environments	No.	Mean	Max	Min	SD	CV (%)	*P* value
BW1	0	Hypoxia	1037	24.8	77.0	3.6	13.4	54.0	0.14
	0	Normoxia	1026	25.4	77.1	2.9	13.1	51.7	
BW2	55	Hypoxia	1037	144.3	328.0	26.0	54.7	37.9	3.81E-07
	56	Normoxia	1026	159.1	394.3	30.2	63.1	39.7	
BW3	104	Hypoxia	907	265.9	498.3	70.5	73.3	27.6	4.17E-08
	105	Normoxia	941	289.4	650.5	63.3	92.5	32.0	
BW4	167	Hypoxia	885	426.4	805.3	117.0	118.9	27.9	2.20E-16
	168	Normoxia	903	533.6	1079.1	68.2	177.2	33.2	
BW5	217	Hypoxia	799	579.6	1003.4	135.5	154.4	26.6	2.20E-16
	218	Normoxia	885	780.9	1588.6	185.7	265.6	34.0	

Body weight (BW), days means the growing out days in either hypoxia or normoxia, the number of animals (No.), maximum (Max), minimum (Min), standard deviation (SD), coefficient of variation (CV%).

Table 2 Phenotypic correlations of body weight across the whole growth period in different environments

Trait	BW1	BW2	BW3	BW4	BW5
BW1	--	0.81	0.61	0.32	0.22
BW2	0.80	--	0.77	0.44	0.32
BW3	0.59	0.80	--	0.66	0.52
BW4	0.29	0.46	0.68	--	0.83
BW5	0.15	0.31	0.56	0.85	--

The spearman's rank correlation coefficient of body weight in hypoxia is presented below diagonal, while the normoxia is above diagonal.

2.2.2 SNP statistic and Population structure

In total 27,090 SNPs that passed SNP minor allele frequency, genotype and individuate call rate criteria, were used for subsequent analysis. Those SNPs were found to be randomly distributed across the genome with a density of approximately 28 SNP per Mb. The highest number of SNPs (4,344) on LG3 while LG11 had the lowest number of SNPs (630) (Figure 1A). A few windows on LG3 show a higher density of SNPs (Figure 1B). Besides this exception, the distribution of SNPs is uniform with the linkage group physical length of the Oreochromis niloticus genome (GenBank accession GCF_001858045).

The PCA represents the genetic structure for individuals from the hypoxic and normoxic environments, respectively (Figures 1C, 1D and Supplementary Figures 3, 4). In the hypoxic environment, the first three principal components (PCs) explain 47.0% of the total genotype-based variation and separate samples according to their family differences. PC1 accounts for 15.2% of the total genotype variation and separates families in hapa3 with other families. In the normoxic environment, the first three components explain 39.8% of the total genotype variation, while the first component accounts for 15.3%. Moreover, the largest PC (PC1) of all samples separates disperse cluster from families in hapa3 again.

These results indicated that there was clear genetic variation caused by family differences in both environments. This was partially caused by the different distribution of the number of fish from four rearing hapas under the normoxic and hypoxic environments. Additionally, the average body weight of fish in hapa3 was larger than that of other hapas, especially the mean body weight of male fish at the first time point was much higher in the normoxic environment than the hypoxic environment (Supplementary Figure 2), indicating that a few families with high body weight dominated in one environment but not the other.

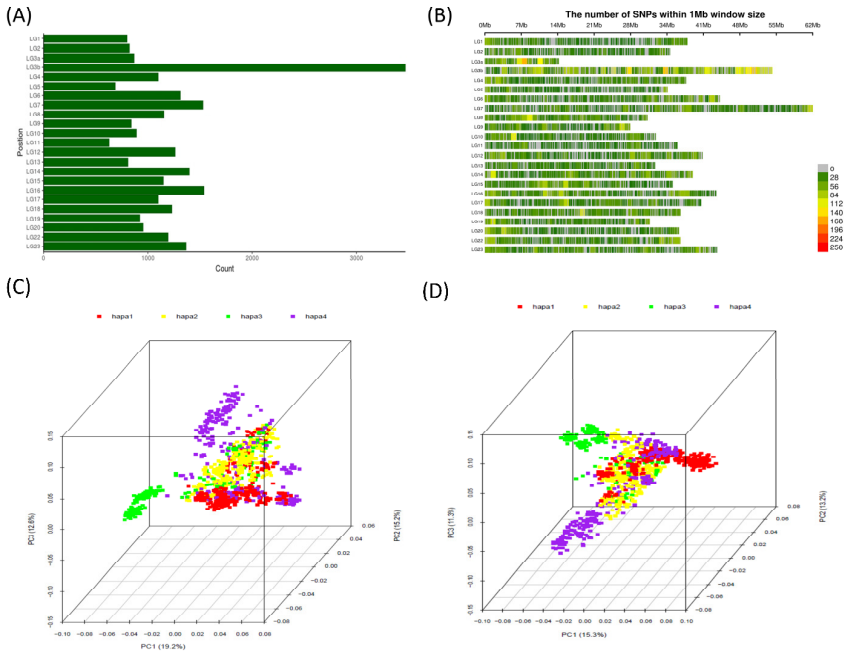


Figure 1. SNP statistics with all individuals. (A) Histogram of SNPs distribution across all linkage groups. (B) SNP density plots across all linkage groups. (C and D) 3D PC plot for origin of tilapia at BW1 in the hypoxic (C) and normoxic (D) environments using all SNPs that passed filtering, where each dot represents one individual.

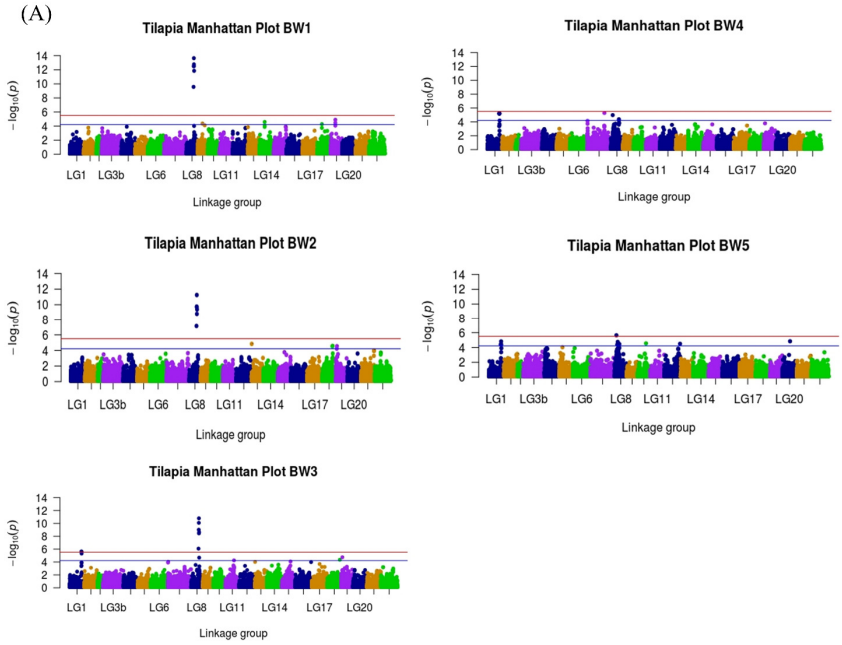
2.2.3 Single environmental GWAS at five different time points

Significant SNPs were detected with a univariate GWAS by implementing a linear mixed model. We observed that sex and hapa effects can explain part of the difference in body weight. Thus, these were treated as fixed factors in our analysis. Overall, five association analyses, one for each time point where body weight was measured, were performed for each environment. The Manhattan plots for each of the five time points in the hypoxic and normoxic environments are shown in Figures 2A and 2B, respectively. In addition, Quantile-Quantile plots with genomic inflation factors were created to aid in estimating the influence of population structure on single environmental GWAS (shown in Supplementary Figures 5 and 6). The P values of corrected thresholds for suggestive and genome-wide significant levels were 4.22 (-log10(1/16504)) and 5.52 (-log10(0.05/16504)), respectively.

In the hypoxic environment, the analyses showed 10 significant and 26 suggestive SNPs associated with BW1 to BW5 (Supplementary Table 2). Among those, six SNPs between 19.48 Mb and 21.04 Mb on LG8 attained genome-wide significance for BW1

to BW3. However, those SNPs were not significant for BW4 and BW5. Two SNPs (LG1: 30766342 and LG1:30766336) were significant associated with BW3 to BW5. Additionally, 16 SNPs above the suggestive level as defined above for BW1 to BW2 were found on LG8, LG18 and LG19, while 18 SNPs mostly located on LG1 and LG8, were found for BW4 to BW5. Interestingly, at BW3, SNPs on LG8 overlapped with BW1 and BW2, while SNPs on LG1 overlapped with BW4 and BW5, further confirming that there is a transition in genomic architecture associated with growth over time.

We also detected 2 significant and 27 suggestive SNPs across different growth stages in the normoxic environment (Supplementary Table 3). The suggestive peak at BW1 covered the same genomic region as that found for the hypoxic environment between 19.48 to 21.03 Mb on LG8. However, similar to the hypoxic environment, the significance of those SNPs declined from BW1 to BW3, a pattern also seen for the SNPs located on LG18 and LG22. A few SNPs on LG7 and LG15 also showed a signal near the suggestive level from BW3 to BW5, which could be potentially interesting, although they did not attain statistical significance.



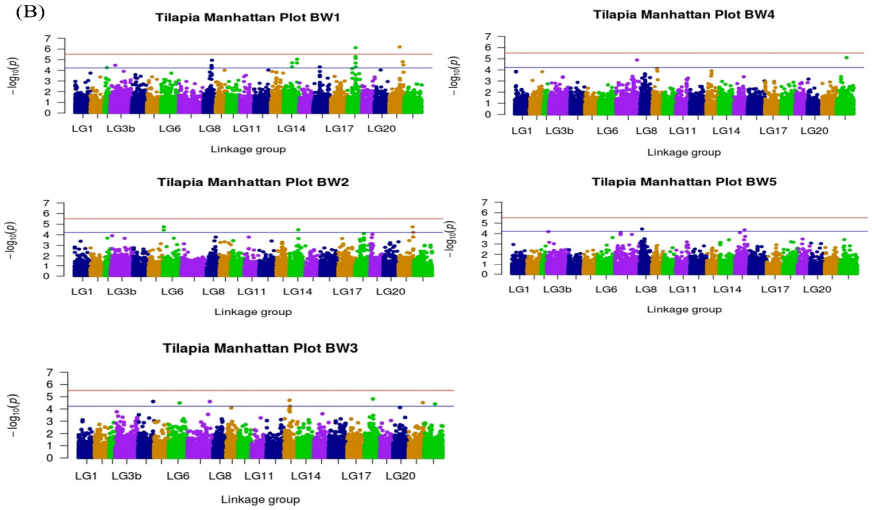


Figure 2. Manhattan plots across the whole growth period in the hypoxic environment (A) and normoxic environment (B). Each dot on this figure corresponds to a SNP within the dataset, while the orange and blue horizontal line represent the genome-wide significance (5.52) and suggestive significance threshold value (4.22), respectively. The Manhattan plots contain –log10 observed P-values for genome-wide SNPs (y-axis) plotted against their corresponding position on each chromosome (x-axis).

2.2.4 Meta-analysis GWAS across two environments

A meta-analysis GWAS that considered the effects of 27,090 SNPs in common in the hypoxic and normoxic environments was performed, and the results are shown in Figure 3. In total 33 SNPs were detected to be significant with five measurements of body weight during the whole growth stage. Clusters of significant SNPs were mostly found on LG8, LG18 and LG22 (Supplementary Table 4). Interestingly, six SNPs located between 19.48 and 21.03 Mb on LG8, three SNPs between 12.44 and 27.32 Mb on LG18 and three SNPs within 1kb at 35.25 Mb on LG22, were all significantly associated with body weight at time points BW1 and BW2. However, the P-values of those SNPs decreased in subsequent growth periods. Five SNPs between 30.54 and 31.19 Mb on LG1, and one SNP on LG15 (LG15:23051993), were associated with body weight from BW3 to BW5. Moreover, two SNPs on LG8 (LG8:4319661, LG8: 11800435) were significant at BW4 and BW5. Notably, those SNPs located on LG8 were found at a different region compared to SNPs on the same LG in hypoxic GWAS. Hence, associations for BW1 to BW2 were different from BW4 to BW5, although BW3 shows both overlap to early and late growth stages, which could indicate that a transition in the pathways involved occurred around this stage.

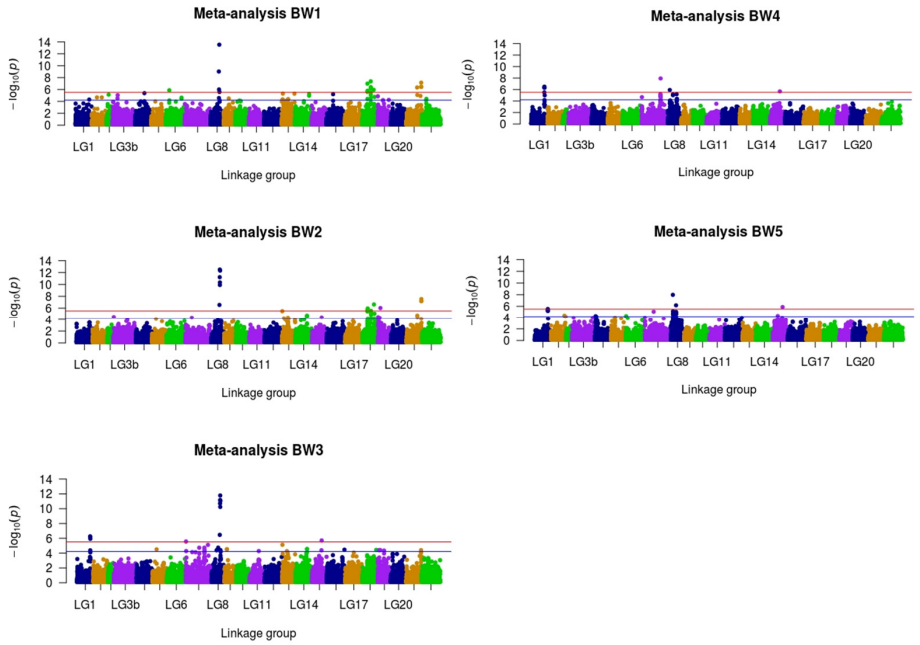


Figure 3. Manhattan plots of Meta-analysis GWAS across two environments. The orange and blue horizontal line represent the genome-wide significance (3.03E-06) and suggestive significance threshold value (6.06E-05), respectively.

2.2.5 Functional annotation analysis

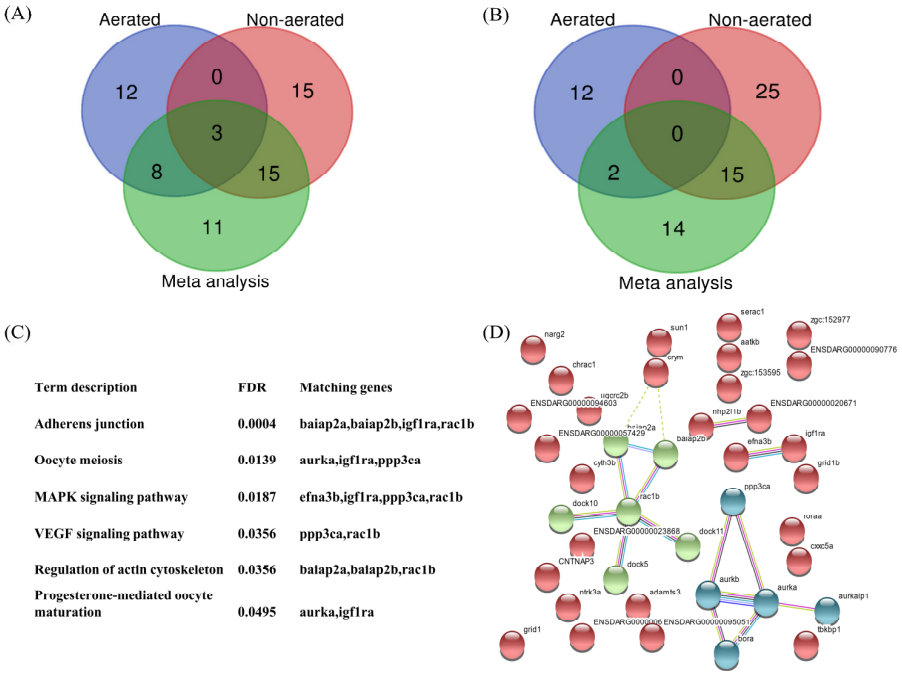
Based on the SNP association patterns for five measurements across the whole growth stage, we defined the early stage as BW1 and BW2, while the later stage is BW3 to BW5. Through gene identification within the associated genomic regions, the functional processes and pathways were subsequently enriched for single environmental and across environmental GWAS, respectively. Considering that BW3 is the transition point, SNPs that overlapped with the early stage were excluded in the functional annotation for the later stage. The candidate genes derived from single environment and across environment GWAS are shown in Figures 4A and 4B, where 15 and 25 genes from the BW1 to BW2 and BW3 to BW5 respectively, were uniquely associated with body weight in the hypoxic environment while another 12 genes were unique to growth in the normoxic environment. It is also noteworthy that three genes (*raraa*, *rarab*, *bahcc1*) were significant for BW1 and BW2 for both single and across environmental GWAS.

During the early growth stage in the hypoxic environment, fourteen GO (Gene ontology) terms were found to be significantly overrepresented (Supplementary table 5), including central nervous system development and steroid hormone

mediated signalling pathways. Six KEGG pathways were found at later growth stage (Figure 4C), including MAPK and VEGF signalling pathways. Protein interaction network analysis showed *dock5*, *dock10*, *dock11*, *baiap2a*, *baiap2b*, *aurka* and *aurkb* strongly interacting with *rac1b* and *ppp3ca,* which all are proteins participating in MAPK and VEGF signalling (Figure 4D).

For the early growth stage of the normoxic environment, retinoic acid receptor signalling pathway, apoptotic signalling pathway, liver development, signal transduction, steroid hormone mediated signalling pathway and brain development biological processes (Supplementary table 6), were significantly enriched, while two (retinoic acid receptor and steroid hormone mediated signalling pathways) overlapped with the same growth period in the hypoxia environment. However, in contrast to the hypoxic environment, we did not find significant terms during the later growth stage in the normoxic environment.

In the meta-analysis GWAS across the normoxic and hypoxic environments, nine GO terms, including retinoic acid receptor signalling pathway and steroid hormone mediated signalling pathway, were mostly enriched in the early growth stage. During the later growth stage, two pathways involved in oocyte meiosis and progesterone-mediated oocyte maturation process. Interestingly, none of hypoxia-related pathway was enriched (Supplementary table 7).



Term description	FDR	Matching genes
Adherens junction	0.0004	baiap2a,baiap2b,igf1ra,rac1b
Oocyte meiosis	0.0139	aurka,igf1ra,ppp3ca
MAPK signaling pathway	0.0187	efna3b,igf1ra,ppp3ca,rac1b
VEGF signaling pathway	0.0356	ppp3ca,rac1b
Regulation of actin cytoskeleton	0.0356	balap2a,balap2b,rac1b
Progesterone-mediated oocyte maturation	0.0495	aurka,igf1ra

Figure 4. Functional annotation based on candidate genomic region associated with growth. (A) Venn diagram summarising the gene count of the early stage (BW1 to BW2) from hypoxia, normoxia and meta-analysis (cross normoxia and hypoxia). (B) Venn diagram summarizing gene count of later stage (BW3 to BW5) from hypoxia, normoxia and meta-analysis. (C) KEGG enrichment of candidate genes in later stage of hypoxia environment (D) protein association network among candidate genes in later stage of the hypoxia environment.

2.3 Discussion

Hypoxia is one of the major environmental factors in fish. Hypoxia tolerance represents the ability of fish species to tolerate low oxygen level and to maintain a sustainable metabolic rate at lower dissolved oxygen levels [20]. Growth is a key trait for aquaculture and can be assessed by weight gain in order to examine the impact of hypoxic condition on fish production. For more than a half century, various and divergent claims have been made regarding the interaction between body size and hypoxia in teleost fish. Recent studies showed that small individuals have the least hypoxia tolerance within some fish species, such as Oscar cichlid [21, 22] and Red seabream [23]. In contrast, small fish chose lower oxygen levels more than large fish in Largemouth bass [24] and Yellow perch [25], however, this behaviour was suggested that the smaller fish utilize the hypoxic zone as refuge protected from the bigger predators [26]. From these studies it is clear that selection for low oxygen is

difficult to ascertain, indicating a clear added value of investigations into genetic consequences of selection, such as the present study.

In general, metabolic rate is highly affected by dissolved oxygen in the rearing environment. Generally, faster growing animals have a higher metabolic rate and therefore require more oxygen. As a consequence, hypoxia is expected to adversely affect fish growth and feed utilization [6]. On the other hand, large individuals have an obvious advantage over small ones in severe hypoxic environments because small fish will use up their glycogen reserves and reach mortality levels much faster with a higher metabolic rate [27]. Overall fish production declines, and disease resistance decreases as a consequence of hypoxia [28]. It has been observed that larger Nile tilapia tolerated low DO levels better than small ones, thought partially due to the fact that Nile tilapia immunity was stronger in larger than smaller [29]. Regardless of the complexity of the relationship between hypoxia and growth, studies focused on the genomic basis of hypoxia-growth interactions in Nile tilapia are sparse.

Our results suggest a number of genes and metabolic pathways involved in the adaptation to differences in dissolved oxygen in Nile tilapia. In the hypoxic environment, 14 significantly enriched processes were associated with the early growth stage, including nervous system development and animal organ development. *Rara* gene codes for the retinoic acid receptor alpha, a transcription factor which regulates genes involved in cellular growth and differentiation [30]. In addition, *raraa* and *rarab* play a key role during development in zebrafish [31]. Mediator of RNA polymerase II transcription subunit 24 (*med24*), an orthologue also found in human, mouse and zebrafish, participates in nervous system development [32]. However, these genes and associated molecular pathways do not indicate a clear link with hypoxia when comparing to other fish studies, and rather might reflect a relation to general growth and development pathways.

During the later growth stage, the results of pathway enrichment suggest that candidate regions are significantly enriched for adherens junctions, oocyte meiosis, MAPK signalling pathway, VEGF signalling pathway, regulation of actin cytoskeleton and progesterone-mediated oocyte maturation. Among these six pathways, various studies in zebrafish, channel catfish, and sea bass have shown MAPK to be involved in low oxygen tolerance in fish [14, 33, 34]. VEGF signalling was shown to be essential for maintaining the vascular density and oxygen supply in tissues [35]. Additionally, the VEGF pathway is also one of the targets of *HIF-1α*, which rapidly accumulates to activate genes involved in a series of responses to hypoxia [8, 36]. The candidate gene *igf1ra,* identified in this study, codes for IGF-1 receptor-a, a receptor of insulin-

like growth factor that was reported to be a primary mediator of growth hormones [37]. The ephrin-A3 gene (*efna3*) is shown as a key functional mediator of hypoxic microenvironment and is regarded as a therapeutic target for hypoxia-specific disease [38]. Retinoic acid receptor-related orphan receptor alpha (*rora*) was demonstrated to be a key regulator of *HIF-1α* activities in human [39]. Finally, the aurora kinase A (*aurka*) gene, a serine kinase in neuroblastoma related to cell growth and migration, can up-regulate expression in human BE(2)-C cells under hypoxia [40]. Recently, Li et al. [16] also found that several regions were significantly related with hypoxia tolerance, including LG3, 4, 11, 14 and 22, especially two regions (LG4:15080000, LG11:24255000) are found to be adjacent with the peak in the hypoxic environment (BW5) of our study. Nevertheless, our results suggest that hypoxia has a non-significant effect on growth during the early growth stage, while, conversely, faster growing tilapia have higher tolerance to hypoxia in the later growing stage, reflected by survival probability. Interestingly, it has been shown that tilapia exposure to a nocturnal hypoxia for 9 weeks led to a better growth performance than normoxia, which is related with a compensatory appetite later in the day [41]. Additionally, Roze et al. [42] has reported that fast growing fish display a better ability to maintain balance to acute hypoxia exposure than slow growing fish, by comparing two genetically different growth strains of Rainbow trout, suggesting a better hypoxia tolerance similar to the findings presented in our study.

In the normoxic environment, six biological processes were significantly enriched for BW1 and BW2, including retinoic acid receptor signalling pathway, apoptotic signalling pathway, liver development, signal transduction, steroid hormone mediated signalling pathway and brain development. Steroid hormone mediated and retinoic acid receptor signalling pathway overlapped with the same stage in the hypoxia environment, which seems mostly involved in general growth and development processes. The overlap in the early growth stage between normoxic and hypoxic environments may result from shared conditions until the first time point. Another possibility is that hypoxia affected small fish less, and there still was sufficient dissolved oxygen as a result of lower overall demand. As fish grew bigger, the metabolic impact of high growth on oxygen consumption and availability may have become more pronounced [43].

For the later growth stage, 12 suggestive SNPs tagging regions containing 22 candidate genes were identified. These included the gene coding for mitochondrial calcium uniporter (*mcu*) that was reported to be a regulator in skeletal muscle growth and homeostasis [44]. The genes coding for oncoprotein-induced transcript 3 (*oit3*) and MAP6 domain containing 1 (*map3d1*) were both reported to be related

with calcium ion binding activity [45]. Yoshida et al. [19] performed the first genome-wide association study to unravel the genetic architecture of harvest weight in a Nile tilapia population derived from a mixture of the 8th generation GIFT and the wild strains from Egypt and Kenya. In that study, four regions were identified that were significantly associated with harvest weight in LG12, 15, 18 and 22, respectively. However, the genes lying in these regions were not significant in our study. One of the reasons could be that the GIFT population has been selected on growth for many generations and those regions have become fixed. This could also explain the limited number of significant SNPs and candidate genes for growth observed in our study. However, it is also likely the specific variants found by Yoshida et al. were never present in our population to begin with.

The results from the meta-analysis show that five genes play a major role in growth and development during the early growth stage, namely *raraa*, *rarab*, *med24*, *brms11a* and *prpf38b*. Two of them (*raraa*, *rarab*) also showed significance for single GWAS in the normoxic and hypoxic environments, respectively. *Prpf38b* only showed a major effect in the hypoxic environment. The orthologues of this gene in human, zebrafish and mouse have been shown to have a function in the central nerve system [46]. Development related genes found in single GWAS, such as *raraa*, *rarab*, and *med24* were significantly associated in the meta-analysis during the later stages. Nucleotide-binding protein 2 (*nump2*) was reported to be associated with both *IGF1* and *IGFP*3 in a human GWA study [47]. Those results suggest that a few major QTLs determine much of the growth rate. Even though growth rate is known to be determined by many genes [48], similarly in human [49] and cattle [50], it was found that a few genes were exceptionally important in explaining genetic variance.

Moreover, no pathway related to hypoxia tolerance was found in meta-analysis GWAS, which indicates some genes affect body weight in the hypoxic environment while different subset of genes is important for body weight under the normoxic environment (see in Figure 4A and 4B). This indicates genotype-by-environment interaction (GxE). However, a GxE analysis for growth rate in the normoxic versus hypoxic environment, based on a quantitative genetic analysis using a genomic relationship matrix derived from the genotyping dataset, showed that the genetic correlation was close to 0.8 [51]. This value suggests some degree of GxE and some reranking of genotypes. Furthermore, there was a large difference in body weight and its variance between environments, which suggests scaling GxE. The genetic correlation of 0.8 suggests that most fish that grow well in a normoxic environment, are also able to grow well in an environment where they experience nocturnal hypoxia. After all, Nile tilapia is a freshwater fish species that has evolved in

environments where hypoxia (e.g. as a result of high temperatures, algal blooms or drought) are nocturnal events. Natural selection would favour animals that would be able to cope with these environments if larger fish would have higher reproductive success.

2.4 Conclusions

Clear associations between genotype and growth were found for both hypoxic and normoxic environments. The associated SNPs, and hence the underlying genomic architecture, however, changed over the growing period. Furthermore, the meta-analysis GWAS across two environments suggested that growth was not under the control by the same genes compared to single environmental GWAS, which we interpret as a genotype-by-environment interaction. The functional annotation confirms that hypoxic stress pathways such as MAPK signalling pathway and VEGF signalling pathway play an important role during the later growth stage in the hypoxic environment. Our findings reveal the genetic complexity of body weight gain under a variety of dissolved oxygen conditions in Nile tilapia, and provide an essential insight into how hypoxia affects body weight gain during the growth stage, which will benefit future tilapia breeding programmes in the context of genomic architecture.

2.5 Methods

2.5.1 Animal Resource

The fish were derived from the Aquaculture Extension Centre of the Malaysian Department of Fisheries at Jitra, Kedah State, Malaysia (6°15’32°N; 100°25’47°E). Genetically Improved Farmed Tilapia (GIFT) strain was used in this experiment, and it had been selected for growth based on estimating breeding value (EBV) of harvest weight, with the genetic gain ranged from 5 to 15 percent per generation. The mate allocation strategy has controlled inbreeding and maintained effective population size [52]. The experimental fish were produced using 72 males from 56 families and 200 females from 73 families (total 81 unique families) of selection line of GIFT generation 15. From each family, fish with EBV for growth that were closed to the family mean EBV were selected as a breeder. The experimental fish were mass produced in four hapas (net-enclosures, each 30m^2) installed in a 500m^2 earthen pond, aerated by a paddlewheel. For each hapa, 18 male and 50 female breeders were distributed for stocking, and they were removed from mating hapa after 15 days. Fry were reared in the same hapas for 60 days until they reached a taggable size. The fingerlings from each rearing hapa were tagged and then transferred into two earthen ponds with an equal number. Overall, 1570 fish were reared in each pond with stocking density of 3 fish/m^2. We managed two ponds with the same

feeding management (i.e. feeding frequency twice per day, feeding rate was adjusted with fish number), while aeration was the only different treatment between two ponds.

We measured DO every two hours for 24hours with a total of 7 days during the different grow-out periods using EcoSense® DO200A. The average DO measurements for aerated pond (normoxia) and non-aerated pond (nocturnal hypoxia) are shown in Supplementary Figure 1 and Supplementary Table 1. Both ponds were normoxic (5mg/L) from 13:00 to 19:00. Non-aerated pond became hypoxic (under 3 mg/L) between 21:00 to next day 9:00. Body weight was measured at five time points (stocking, 55/56 days, 104/105 days, 167/168 days and 217/218 days) growing out in the hypoxic and normoxic environments, respectively. Fish were euthanized using clove oil at a dose of 400 ppm after the experiment. Fin clips were preserved in 95% ethanol and stored at −20 °C until DNA extraction. More details about this experiment can be found in Mengistu et al. [51].

2.5.2 Genotyping, variant calling and quality control

DNA extraction and genotyping procedures were described in previous study by Mengistu et al. [51]. In short, we isolated genomic DNA from tilapia fin clips using the DNeasy Blood and Tissue kit. DNA quality was assessed by 260/280 and 260/230 ratios on NanoDrop 2000 spectrophotometer. DNA concentration was measured with Qubit 2.0 Fluorometer. DNA samples were digested with ApeKI, and polymerase chain reaction (PCR) was used to amplify fragments varied from 170 to 350 bp. The prepared libraries were sequenced on the Illumina HiSeq 2000 platform.

Raw sequence reads were trimmed for adaptors and low-quality bases with Sickle (https://github.com/najoshi/sickle). The quality of each individual was evaluated by FastQC (version 1.6) [53]. Sequence mapping for 2171 individuals was performed using bwa -mem algorithm [54] aligning to the tilapia reference genome (GenBank accession GCF_001858045.1). Variant calling was analysed with FreeBayes (version 1.0.2) [55] in a default setting except for these parameters: --min-base-quality 10, --haplotype-length 0 and --ploidy 2. The SNP data was further filtered by Plink (version1.9) [56] with the following exclusion criteria: Minor Allele Frequency < 2%, genotyping call-rate for SNPs < 80% and individual rate < 70%. Finally, a total of 2063 individuals and 27,090 SNPs were used for subsequent analyses.

2.5.3 Statistic description, Population Structure and Association Analysis

Basic statistics of phenotype data was analysed in R (version 3.5.3). Body weight in our study is not completely following a normal distribution as estimated by Shapiro-Wilk test [57]. Therefore, we compare two paired groups at five time point using the Wilcoxon test. The phenotypic correlation was calculated by spearman's rank correlation coefficient method. Then, body weight was transformed to better fit the normal distribution by square root method [58]. To estimate the influence of factors such as hapa (early rearing environment) and sex in our experiment, they were tested in a linear model using Stepwise Algorithm [59] with the formula: $\boldsymbol{y_{ij}} = \boldsymbol{u} + \boldsymbol{\alpha_i} + \boldsymbol{\beta_j} + \boldsymbol{\alpha_i} * \boldsymbol{\beta_j} + \boldsymbol{\varepsilon_{ij}}$, while $\boldsymbol{y}$ is the body weight; $\boldsymbol{u}$ is the population mean; $\boldsymbol{\alpha_i}$ is the effect of the ith level of hapa; $\boldsymbol{\beta_j}$ is the effect of the jth level of sex; ε is the random error effect. It suggested that hapa, sex and their interaction were significant with body weight. Therefore, residuals from the fixed effects model were used for the subsequent association analysis [60].

A principal component analysis (PCA) was performed to estimate population structure before GWAS in Plink (version 1.9) [56]. The top five principal components were added as covariates and included in subsequent GWAS model as fixed effect to account for the sample structure in this association analysis. Considering the Bonferroni method being overly conservative, we defined the genome-wide significant using the SimpleM method [61]. In total 16,504 independent tests were calculated based on LD (linkage disequilibrium) characteristics. The significant and suggestive lines are 1 and 5% genome-wide significant divided by the SNPs number of independent SNPs in the association. Given the number of effective independent tests, the thresholds for genome-wide and suggestive significance P-value were evaluated as 3.03E-06 (0.05/16504) and 6.06E-05 (1/16504), respectively.

A univariate GWAS was performed by implementing a linear mixed model in GEMMA [62]:

$$\boldsymbol{y} = \boldsymbol{W\alpha} + \boldsymbol{x\beta} + \boldsymbol{\mu} + \boldsymbol{\varepsilon}$$

In this equation, **y** is a vector of observation on body weight; $\boldsymbol{W}$ is a covariate matrix of fixed effects (including top five PCs) used to adjust population structure; $\boldsymbol{\alpha}$ is a vector of the corresponding coefficient including the intercept; $\boldsymbol{x}$ is a vector of the marker genotypes and $\boldsymbol{\beta}$ is the corresponding vector of marker effects for the phenotypes; $\boldsymbol{\mu}$ is a vector of random effects and ε is the random residuals. We performed the Wald statistic for each SNP which means we tested the alternative

hypothesis H_1: $\boldsymbol{\beta} \neq 0$ compared to null hypothesis H_0: $\boldsymbol{\beta} = 0$ for each SNP, which is one of common methods in GWAS studies of quantitative traits [63] .

Meta-analysis is powerful to detect shared genetic architecture across traits and populations [64]. Thus, we applied an inverse-variance weighted (IVW) method to estimate the SNP effect and significance combined normoxic and hypoxic environments through Meta (Version 1.7) [65, 66] . The weight (w_i) for ith environment was calculated by the following equation:

$$\mathbf{w_i} = \frac{\mathbf{1}}{\mathbf{s_i^2}}$$

Here s_i is the standard error of the SNP effect in i^{th} environment GWAS. Then, the effect size and standard error for i^{th} environment GWAS were estimated by the following:

$$\boldsymbol{\beta} = \frac{\sum_{i=1}^{2} \mathbf{w_i}\boldsymbol{\beta_i}}{\sum_{i=1}^{2} \mathbf{w_i}}$$

$$\mathbf{s^2} = \frac{\mathbf{1}}{\sum_{i=1}^{2} \mathbf{w_i}}$$

The statistical significance was estimated by a z-score of IVW as bellow:

$$\mathbf{Z} = \frac{\boldsymbol{\beta}}{\mathbf{s}} = \frac{\sum_{i=1}^{2} \mathbf{w_i}\boldsymbol{\beta_i}}{\sqrt{\sum_{i=1}^{2} \mathbf{w_i}}}$$

2.5.4 Post-GWAS analysis

Manhattan and quantile-quantile (Q-Q) plots were generated through the “qqman” package (https://cran.r-project.org/web/ packages/qqman/). The inflation factor λ was calculated to indicate the influence of population structure in the association analyses. Candidate regions were defined as the genomic regions that located 20 kb upstream and downstream of the genome-wide significant SNPs. In order to identify candidate genes nearby the significant SNPs, we used the Custom Annotations function to create an annotation set with parameters (--distance 20000 --gene_phenotype --symbol) in Ensembl Variant Effect Predictor (VEP) [67]. All protein sequences of candidate genes were extracted through reference protein sequence with an inhouse python script, and were further used for functional enrichment analysis in STRING V11.0 [68]. The false discovery rate (FDR) adjusted p-value of 0.05 was used to define significant enrichment.

2.6 Abbreviations

AMPK: AMP-activated protein kinase; AURKA: aurora kinase A; BW: body weight; CV: coefficient of variation; DO: dissolved oxygen; EBV: estimated breeding value; EFNA3: ephrin-A3; FDR: false discovery rate; GIFT: Genetically Improved Farmed Tilapia; GO: Gene ontology; GWAS: Genome Wide Association Study; GxE: genotype-by-environment interaction; HAS1: Hyaluronan synthase; HIF: hypoxia-Inducible factors; IVW: inverse-variance weighted; KEGG: Kyoto Encyclopedia of Genes and Genomes; LD: linkage disequilibrium; LG: lineage group; MAPK: mitogen-activated protein kinase; MAP3D1: MAP6 domain containing 1; MCU: mitochondrial calcium uniporter; MED24: Mediator of RNA polymerase II transcription subunit 24; NUMP2: nucleotide-binding protein 2; OIT3: oncoprotein-induced transcript 3; PCA: principal component analysis; PC: principal component; PGK1: phosphoglycerate kinase 1; Q-Q: quantile-quantile; RORA: retinoic acid receptor-related orphan receptor alpha; ROS: reactive oxygen species; SD: standard deviation; SNP: single nucleotide polymorphism; TH: tyrosine hydroxylase; VEGF: vascular endothelial growth factor; VEP: Variant Effect Predictor.

2.7 Acknowledgements

The authors gratefully acknowledge WorldFish breeding team for their field and technical assistance. We also thank all colleagues who provided help during analysis procedures from Animal Breeding and Genomics group.

2.8 Authors' contributions

JAHB, HAM, HK and MAMG conceived and designed the experiments. SBM and HAM contributed to animal experiment and collected biological samples, and they also supported data analysis and result interpretation. XFY, HJM wrote the manuscript. XFY, JWMB and HJM contributed to data analysis. XFY, HJM, MAMG and HK contributed to result interpretation. SBM, JWMB, HAM, JAHB, MAMG and HK contributed to manuscript revision. All authors reviewed and approved the final manuscript.

2.9 Funding

This work was funded by the CGIAR Research Program on Fish Agri-Food Systems (FISH) led by WorldFish. The original experiment was funded by the Koepon Foundation with additional support to WorldFish by the International Fund for Agricultural Development (IFAD). The authors also thank to the financial support from China Scholarship Council (CSC, File No.201803250069), which has sponsored Xiaofei Yu's PhD study at Wageningen University and Research. The funding bodies

played no role in design of the study and data collection, analysis, and result interpretation in the manuscript.

2.10 Availability of data and materials

The genotype and phenotype data generated or analysed during this study are in the Harvard Dataverse repository with accession number KCBEON, which can be accessed at https://doi.org/10.7910/DVN/KCBEON. The tilapia reference genome (GCF_001858045.1_ASM185804v2_genomic.fna.gz) and annotation file (GCF_001858045.1_ASM185804v2_genomic.gff.gz) were downloaded from the NCBI genome assembly website (https://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/001/858/045/GCF_001858045.1_ASM185804v2/). The authors declare that all data supporting the findings are available within this article and its supplementary files.

2.11 Declarations

2.11.1 Ethics approval and consent to participate

This study was approved by the internal WorldFish ethics committee. All the parties agreed for this experiment.

2.11.2 Competing interests

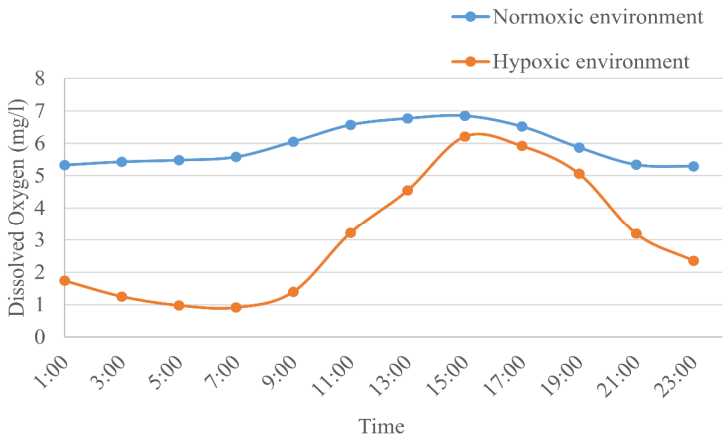
MAMG is a member of the editorial board of BMC Genomics journal. The authors declare that they have no competing interests.

2.11.3 Consent for publication

Not applicable.

2.12 Supplementary Information

For a compact layout, I have not included all Supplementary material in the thesis, only those which may assist the reader. For more information, the complete Supplementary figures and tables of the article are available at https://doi.org/10.1186/s12864-021-07486-5.



Supplementary Figure 1. Variation of dissolved oxygen in the normoxic and hypoxic environments during the 24-hour cycle.

Supplementary Figure 2. Body weight comparison amongst four hapas in the normoxic and hypoxic environments.

Supplementary Figure 3. Two-dimensional plots of all individuals using SNP markers in the hypoxic environment.

Supplementary Figure 4. Two-dimensional plots of all individuals using SNP markers in the normoxic environment.

Supplementary Figure 5. Quantile-quantile plots through the whole growth stage in the hypoxic environment.

Supplementary Figure 6. Quantile-quantile plots through the whole growth stage in the normoxic environment.

Supplementary Table 1. Measurements of dissolved oxygen in the normoxic and hypoxic environments.

Supplementary Table 2. Information of genome-wide significant and suggestive SNPs in the hypoxic environment.

References

1. FAO: The State of World Fisheries and Aquaculture 2020. Rome: Sustainability in action.
2. Guyon R, Rakotomanga M, Azzouzi N, Coutanceau JP, Bonillo C, D'Cotta H, Pepey E, Soler L, Rodier-Goud M, D'Hont A *et al*: A high-resolution map of the Nile tilapia genome: a resource for studying cichlids and other percomorphs. *BMC Genomics.* 2012;13:222.
3. Ponzoni RW, Nguyen NH, Khaw HL, Hamzah A, Bakar KRA, Yee HY: Genetic improvement of Nile tilapia (*Oreochromis niloticus*) with special reference to the work conducted by the WorldFish Center with the GIFT strain. *Reviews in Aquaculture.* 2011;3(1):27-41.
4. Portner HO: Oxygen- and capacity-limitation of thermal tolerance: a matrix for integrating climate-related stressor effects in marine ecosystems. *J Exp Biol.* 2010;213(6):881-893.
5. Fitzgibbon QP, Strawbridge A, Seymour RS: Metabolic scope, swimming performance and the effects of hypoxia in the mulloway, Argyrosomus japonicus (Pisces : Sciaenidae). *Aquaculture.* 2007;270(1-4):358-368.
6. Mengistu SB, Mulder HA, Benzie JAH, Komen H: A systematic literature review of the major factors causing yield gap by affecting growth, feed conversion ratio and survival in Nile tilapia (*Oreochromis niloticus*). *Reviews in Aquaculture.* 2020;12(2):524-541.
7. Michiels C: Physiological and pathological responses to hypoxia. *Am J Pathol.* 2004;164(6):1875-1882.
8. Majmundar AJ, Wong WHJ, Simon MC: Hypoxia-Inducible Factors and the Response to Hypoxic Stress. *Mol Cell.* 2010;40(2):294-309.
9. Nikinmaa M, Rees BB: Oxygen-dependent gene expression in fishes. *American journal of physiology Regulatory, integrative and comparative physiology.* 2005;288(5):R1079-1090.
10. Xiao W: The hypoxia signaling pathway and hypoxic adaptation in fishes. *Sci China Life Sci.* 2015;58(2):148-155.
11. Zhu CD, Wang ZH, Yan BA: Strategies for hypoxia adaptation in fish species: a review. *Journal of Comparative Physiology B-Biochemical Systems and Environmental Physiology.* 2013;183(8):1005-1013.
12. Du SNN, Mahalingam S, Borowiec BG, Scott GR: Mitochondrial physiology and reactive oxygen species production are altered by hypoxia acclimation in killifish (*Fundulus heteroclitus*). *J Exp Biol.* 2016;219(8):1130-1138.
13. Zhong X, Wang X, Zhou T, Jin Y, Tan S, Jiang C, Geng X, Li N, Shi H, Zeng Q *et al*: Genome-Wide Association Study Reveals Multiple Novel QTL Associated with Low Oxygen Tolerance in Hybrid Catfish. *Marine biotechnology.* 2017;19(4):379-390.
14. Wang XZ, Liu SK, Jiang C, Geng X, Zhou T, Li N, Bao LS, Li Y, Yao J, Yang YJ *et al*: Multiple across-strain and within-strain QTLs suggest highly complex genetic architecture for hypoxia tolerance in channel catfish. Molecular Genetics and Genomics. 2017;292(1):63-76
15. Anttila K, Dhillon RS, Boulding EG, Farrell AP, Glebe BD, Elliott JAK, Wolters WR, Schulte PM: Variation in temperature tolerance among families of Atlantic salmon (*Salmo salar*) is associated with hypoxia tolerance, ventricle size and myoglobin level. *J Exp Biol.* 2013;216(7):1183-1190.
16. Li HL, Gu XH, Li BJ, Chen CH, Lin HR, Xia JH: Genome-Wide QTL Analysis Identified Significant Associations Between Hypoxia Tolerance and Mutations in the GPR132 and ABCG4 Genes in Nile Tilapia. *Marine biotechnology.* 2017;19(5):441-453.

17. Aquaculture Genomics G, Breeding W, Abdelrahman H, ElHady M, Alcivar-Warren A, Allen S, Al-Tobasei R, Bao L, Beck B, Blackburn H *et al*: Aquaculture genomics, genetics and breeding in the United States: current status, challenges, and priorities for future research. *BMC Genomics*. 2017;18(1):191.
18. Brennan RS, Healy TM, Bryant HJ, Van La M, Schulte PM, Whitehead A: Integrative Population and Physiological Genomics Reveals Mechanisms of Adaptation in Killifish. *Molecular biology and evolution.* 2018;35(11):2639-2653.
19. Yoshida GM, Lhorente JP, Correa K, Soto J, Salas D, Yanez JM: Genome-Wide Association Study and Cost-Efficient Genomic Predictions for Growth and Fillet Yield in Nile Tilapia (*Oreochromis niloticus*). *G3.* 2019.
20. Mandic M, Todgham AE, Richards JG: Mechanisms and evolution of hypoxia tolerance in fish. *Proc Biol Sci.* 2009;276(1657):735-744.
21. Almeida-Val VM, Val AL, Duncan WP, Souza FC, Paula-Silva MN, Land S: Scaling effects on hypoxia tolerance in the Amazon fish *Astronotus ocellatus* (Perciformes: Cichlidae): contribution of tissue enzyme levels. *Comp Biochem Physiol B Biochem Mol Biol.* 2000;125(2):219-226.
22. Sloman KA, Wood CM, Scott GR, Wood S, Kajimura M, Johannsson OE, Almeida-Val VM, Val AL: Tribute to R. G. Boutilier: the effect of size on the physiological and behavioural responses of oscar, *Astronotus ocellatus*, to hypoxia. *J Exp Biol.* 2006;209(Pt 7):1197-1205.
23. Ishibashi Y, Inoue K, Nakatsukasa H, Ishitani Y, Miyashita S, Murata O: Ontogeny of tolerance to hypoxia and oxygen consumption of larval and juvenile red sea bream, *Pagrus major*. *Aquaculture.* 2005;244(1-4):331-340.
24. Burleson ML, Wilhelm DR, Smatresk NJ: The influence of fish size on the avoidance of hypoxia and oxygen selection by largemouth bass. *J Fish Biol.* 2001;59(5):1336-1349.
25. Robb T, Abrahams MV: Variation in tolerance to hypoxia in a predator and prey species: an ecological advantage of being small? *J Fish Biol.* 2003;62(5):1067-1081.
26. Vejrik L, Matejickova I, Juza T, Frouzova J, Seda J, Blabolil P, Ricard D, Vasek M, Kubecka J, Riha M *et al*: Small fish use the hypoxic pelagic zone as a refuge from predators. *Freshwater Biol.* 2016;61(6):899-913.
27. Nilsson GE, Ostlund-Nilsson S: Does size matter for hypoxia tolerance in fish? *Biol Rev.* 2008;83(2):173-189.
28. Abdel-Tawwab M, Monier MN, Hoseinifar SH, Faggio C: Fish response to hypoxia stress: growth, physiological, and immunological biomarkers. *Fish Physiol Biochem.* 2019;45(3):997-1013.
29. Abdel-Tawwab M, Hagras AE, Elbaghdady HAM, Monier MN: Effects of dissolved oxygen and fish size on Nile tilapia, *Oreochromis niloticus* (L.): growth performance, whole-body composition, and innate immunity. *Aquacult Int.* 2015;23(5):1261-1274.
30. Wang XJ, Dasari S, Nowakowski GS, Lazaridis KN, Wieben ED, Kadin ME, Feldman AL, Boddicker RL: Retinoic acid receptor alpha drives cell cycle progression and is associated with increased sensitivity to retinoids in T-cell lymphoma. *Oncotarget.* 2017;8(16):26245-26255.
31. Hale LA, Tallafuss A, Yan YL, Dudley L, Eisen JS, Postlethwait JH: Characterization of the retinoic acid receptor genes raraa, rarab and rarg during zebrafish development. *Gene Expr Patterns.* 2006;6(5):546-555.
32. Harrison C, Wabbersen T, Shepherd IT: In Vivo Visualization of the Development of the Enteric Nervous System Using a Tg(-8.3bphox2b:Kaede) Transgenic Zebrafish. Genesis. 2014;52(12):985-990.
33. Tian Y, Wen HS, Qi X, Zhang XY, Li Y: Identification of mapk gene family in Lateolabrax maculatus and their expression profiles in response to hypoxia and salinity challenges. *Gene.* 2019;684:20-29.

34. Marques IJ, Leito JTD, Spaink HP, Testerink J, Jaspers RT, Witte F, van Den Berg S, Bagowski CP: Transcriptome analysis of the response to chronic constant hypoxia in zebrafish hearts. *Journal of Comparative Physiology B-Biochemical Systemic and Environmental Physiology.* 2008;178(1):77-92.
35. Maitland ML, Lou XJ, Ramirez J, Desai AA, Berlin DS, McLeod HL, Weichselbaum RR, Ratain MJ, Altman RB, Klein TE: Vascular endothelial growth factor pathway. *Pharmacogenet Genom.* 2010;20(5):346-349.
36. Kurihara T, Westenskow PD, Friedlander M: Hypoxia-Inducible Factor (HIF)/Vascular Endothelial Growth Factor (VEGF) Signaling in the Retina. *Adv Exp Med Biol.* 2014;801:275-281.
37. Hakuno F, Takahashi SI: IGF1 receptor signaling pathways. *J Mol Endocrinol.* 2018;61(1):T69-T86.
38. Husain A, Chiu YT, Ho DW, Sze KM, Chan LK, Tsui YM, Wong CC, Ng IO: EFNA3, a key functional mediator of hypoxic microenvironment in hepatocellular carcinoma. *Cancer Research.* 2018;78(13).
39. Li H, Zhou L, Dai J: Retinoic acid receptor-related orphan receptor RORalpha regulates differentiation and survival of keratinocytes during hypoxia. *J Cell Physiol.* 2018;233(1):641-650.
40. Romain CV, Paul P, Lee S, Qiao J, Chung DH: Targeting aurora kinase A inhibits hypoxia-mediated neuroblastoma cell tumorigenesis. *Anticancer Res.* 2014;34(5):2269-2274.
41. Obirikorang KA, Acheampong JN, Duodu CP, Skov PV: Growth, metabolism and respiration in Nile tilapia (*Oreochromis niloticus*) exposed to chronic or periodic hypoxia. *Comp Biochem Physiol A Mol Integr Physiol.* 2020;248:110768.
42. Roze T, Christen F, Amerand A, Claireaux G: Trade-off between thermal sensitivity, hypoxia tolerance and growth in fish. *J Therm Biol.* 2013;38(2):98-106.
43. Mandic M, Regan MD: Can variation among hypoxic environments explain why different fish species use different hypoxic survival strategies? *J Exp Biol.* 2018;221(21).
44. Kwong JQ, Huo JZ, Bround MJ, Boyer JG, Schwanekamp JA, Ghazal N, Maxwell JT, Jang YC, Khuchua Z, Shi K *et al*: The mitochondrial calcium uniporter underlies metabolic fuel preference in skeletal muscle. *Jci Insight.* 2018;3(22).
45. Sprague J, Bayraktaroglu L, Clements D, Conlin T, Fashena D, Frazer K, Haendel M, Howe DG, Mani P, Ramachandran S *et al*: The Zebrafish Information Network: the zebrafish model organism database. *Nucleic acids research.* 2006;34(Database issue):D581-585.
46. Koscielny G, Yaikhom G, Iyer V, Meehan TF, Morgan H, Atienza-Herrero J, Blake A, Chen CK, Easty R, Di Fenza A *et al*: The International Mouse Phenotyping Consortium Web Portal, a unified point of access for knockout mice and related phenotyping data. *Nucleic acids research.* 2014;42(Database issue):D802-809.
47. Teumer A, Qi Q, Nethander M, Aschard H, Bandinelli S, Beekman M, Berndt SI, Bidlingmaier M, Broer L, Group CLW *et al*: Genomewide meta-analysis identifies loci associated with IGF-I and IGFBP-3 levels with impact on age-related traits. *Aging Cell.* 2016;15(5):811-824.
48. Wood AR, Esko T, Yang J, Vedantam S, Pers TH, Gustafsson S, Chun AY, Estrada K, Luan J, Kutalik Z *et al*: Defining the role of common variation in the genomic and biological architecture of adult human height. *Nature genetics.* 2014;46(11):1173-1186.
49. Couto Alves A, De Silva NMG, Karhunen V, Sovio U, Das S, Taal HR, Warrington NM, Lewin AM, Kaakinen M, Cousminer DL *et al*: GWAS on longitudinal growth traits reveals different genetic factors influencing infant, child, and adult BMI. *Sci Adv.* 2019;5(9):eaaw3095.
50. Seabury CM, Oldeschulte DL, Saatchi M, Beever JE, Decker JE, Halley YA, Bhattarai EK, Molaei M, Freetly HC, Hansen SL *et al*: Genome-wide association study for feed efficiency and growth traits in U.S. beef cattle. *BMC Genomics.* 2017;18(1):386.

51. Mengistu SB, Mulder HA, Benzie JA, Khaw HL, Megens H-J, Trinh TQ, Komen H: Genotype by environment interaction between aerated and non-aerated ponds and the impact of aeration on genetic parameters in Nile tilapia (*Oreochromis niloticus*). *Aquaculture*. 2020:735704.
52. Ponzoni RW, Khaw HL, Nguyen NH, Hamzah A: Inbreeding and effective population size in the Malaysian nucleus of the GIFT strain of Nile tilapia (*Oreochromis niloticus*). *Aquaculture.* 2010;302(1-2):42-48.
53. Andrews S: FastQC: a quality control tool for high throughput sequence data. In.: Babraham Bioinformatics, Babraham Institute, Cambridge, United Kingdom; 2010.
54. Li H, Durbin R: Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics.* 2009;25(14):1754-1760.
55. Garrison E, Marth G: Haplotype-based variant detection from short-read sequencing. *arXiv preprint arXiv:12073907.* 2012.
56. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, Maller J, Sklar P, de Bakker PI, Daly MJ *et al*: PLINK: a tool set for whole-genome association and population-based linkage analyses. *American journal of human genetics.* 2007;81(3):559-575.
57. Villasenor Alva JA, Estrada EG: A generalization of Shapiro–Wilk's test for multivariate normality. *Communications in Statistics—Theory and Methods.* 2009;38(11):1870-1883.
58. McDonald JH: Handbook of biological statistics, vol. 2: sparky house publishing Baltimore, MD; 2009.
59. Neerchal NK, Morel J, Huang X, Moluh A: A stepwise algorithm for generalized linear mixed models. In: *SAS Global Forum, Washington, DC: 2014*; 2014: 1822-2014.
60. Gondro C, Van der Werf J, Hayes B: Genome-wide association studies and genomic prediction: Springer; 2013.
61. Gao XY, Becker LC, Becker DM, Starmer JD, Province MA: Avoiding the High Bonferroni Penalty in Genome-Wide Association Studies. *Genetic epidemiology.* 2010;34(1):100-105.
62. Zhou X, Stephens M: Genome-wide efficient mixed-model analysis for association studies. *Nature genetics.* 2012;44(7):821-U136.
63. Wu B, Pankow JS: Fast and Accurate Genome-Wide Association Test of Multiple Quantitative Traits. *Comput Math Methods Med*. 2018;2018:2564531.
64. Li YR, Li J, Zhao SD, Bradfield JP, Mentch FD, Maggadottir SM, Hou C, Abrams DJ, Chang D, Gao F *et al*: Meta-analysis of shared genetic architecture across ten pediatric autoimmune diseases. *Nat Med*. 2015;21(9):1018-1027.
65. Liu JZ, Tozzi F, Waterworth DM, Pillai SG, Muglia P, Middleton L, Berrettini W, Knouff CW, Yuan X, Waeber G *et al*: Meta-analysis and imputation refines the association of 15q25 with smoking quantity. *Nature genetics.* 2010;42(5):436-U475.
66. Lee CH, Cook S, Lee JS, Han B: Comparison of Two Meta-Analysis Methods: Inverse-Variance-Weighted Average and Weighted Sum of Z-Scores. *Genomics Inform.* 2016;14(4):173-180.
67. McLaren W, Gil L, Hunt SE, Riat HS, Ritchie GRS, Thormann A, Flicek P, Cunningham F: The Ensembl Variant Effect Predictor. *Genome Biol.* 2016;17.
68. Szklarczyk D, Gable AL, Lyon D, Junge A, Wyder S, Huerta-Cepas J, Simonovic M, Doncheva NT, Morris JH, Bork P *et al*: STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic acids research.* 2019;47(D1):D607-D613.

3

Quantitative trait loci controlling swimming performance and their effect on growth in Nile tilapia (*Oreochromis niloticus*)

Xiaofei Yu[1], Samuel Bekele Mengistu[1, 2], Han A. Mulder[1], Arjan P. Palstra[1], John A. H. Benzie[3, 4], Trong Quoc Trinh[3], Martien A.M. Groenen[1], Hans Komen[1], Hendrik-Jan Megens[1]

[1] Animal Breeding and Genomics, Wageningen University & Research, The Netherlands

[2]School of Animal and Range Sciences, College of Agriculture, Hawassa University, Hawassa Ethiopia

[3]WorldFish Centre, Jalan Batu Maung, Bayan Lepas, Penang, Malaysia

[4]School of Biological Earth and Environmental Sciences, University College Cork, Cork, Ireland

Aquaculture 560 (2022): 738522.

Abstract

Critical swimming speed (U_{crit}) is an important measurement of swimming performance and a good indicator for cardio-respiratory health. It offers a new opportunity to select fish with better fitness. However, the genomic architecture of swimming performance at whole genome level is not clear in Nile tilapia. For this study, swimming performance was measured in 1500 fish from the Genetic Improvement of Farmed Tilapia strain in their early life, which were subsequently grown in a non-aerated pond (nocturnal hypoxia) until harvest. Our results showed that the heritability for U_{crit} was 0.31 ± 0.04. Genetic correlations between U_{crit} and harvest weight (−0.13 ± 0.13) and between U_{crit} and daily growth coefficient (DGC) (−0.26 ± 0.13) were slightly negative. Nine single nucleotide polymorphisms (SNPs) were found to be suggestively associated with U_{crit}, of which five are located in a region between 12.18 and 19.89 Mb on linkage group (LG)14, while two SNPs are located between 18.85 Mb to 18.94 Mb on LG13. The remaining two SNPs are located on LG19 and LG12, respectively. Candidate genes in high linkage disequilibrium (LD) with these SNPs were identified, including *hip1*, *hectd1*, *elna*, *smyd1b*, *rrp12* and *pprc1*. This suggests possible involvement of neuronal growth, muscle activity, cardiovascular development and angiogenesis, and oxygen/hypoxia regulation. Three of these nine SNPs were significantly associated with both harvest weight and DGC, and SNP genotypes that associated with lowest mean U_{crit} were associated with highest mean harvest weight and DGC. In conclusion, we found a clear pleiotropic effect of some SNPs that affect both growth and swimming performance in a hypoxic environment, while other SNPs had only effect on swimming performance, but not on growth. Although fast swimming fish are assumed to show slower growth, such as lower DGC and harvest weight, candidate genetic markers identified in this study provide an opportunity to select fish with good cardio-respiratory health and growth.

Keywords: Nile tilapia, hypoxia, swimming performance, growth, QTL

3.1 Introduction

Tilapia is the second most important farmed fish species for aquaculture production. It is farmed in >120 countries across the world in a wide range of culture environments. Global tilapia aquaculture production grew 11% annually and increased from 383,654 metric tons in 1990 to almost 7,000,000 metric tons in 2020 according to FAO Fisheries and Aquaculture statistics (El-Sayed, 2019) and the worldwide tilapia market was US$ 7.9 billion in 2020 (from IMARC, https://www.imarcgroup.com/tilapia-market). Many selective breeding programs have been established (Tayamen, 2004; Ponzoni et al., 2011; Thodesen et al., 2011). One of the most important breeding programs is called the "Genetic Improvement of Farmed Tilapia" (GIFT) implemented by WorldFish in Malaysia, which has played an important role in boosting tilapia production in many countries in Asia and the Pacific region (Ponzoni et al., 2011; Bentsen et al., 2017). However, a yield gap is often observed in environments where fish are farmed without aeration. Low levels of dissolved oxygen, can adversely affect growth, feed conversion ratio and survival in Nile tilapia (Mengistu et al., 2020b). For most smallholder farmers, aeration of fishponds is not available or too expensive. In non-aerated ponds extreme hypoxia (below the critical level of 3 mg/L) can frequently occur, especially during the night when algae become net oxygen consumers (Mengistu et al., 2020a; Yu et al., 2021). Therefore, it is crucial to select tilapia that grow better and healthier under conditions where dissolved oxygen is limited (such as hypoxia) by breeding companies and organizations. Smallholder tilapia farmers can then use these for more efficient fingerling production.

Swimming performance of fishes has been widely studied over half a century (Kieffer, 2010). Swimming performance is an important feature that correlates with fitness, survival and metabolism in aquacultural species (Palstra and Planas, 2012). It also plays a crucial role in several other aspects, such as migration, habitat selection, predator-prey interaction, and reproduction (Cano-Barbacil et al., 2020). Prolonged swimming performance in fish can be measured in a critical swimming challenge test, by which critical swimming speed (U_{crit}) can be assessed. During the test, swimming speeds are incrementally increased at prescribed intervals until fish become fatigued and stop swimming. The moment of fatigue determines the U_{crit} which is primarily aerobically driven and at which maximum oxygen uptake occurs (as smoothened average over this last swimming period but with significantly higher peaks during burst-and-glide swimming (Brett, 1964; Plaut, 2001; Kieffer, 2010; Palstra and Planas, 2012; Zhang et al., 2019). In general, U_{crit} is used as indicator to evaluate aerobic swimming performance and physical fitness in fish, similar to using a treadmill for human or rodents. For marine fish species such as Atlantic salmon

(*Salmo sala*) and Gilthead Seabream (*Sparus aurata*), U_{crit} was used as a predictor for target traits such as growth performance in relation to feed intake and fillet percentage (Palstra et al., 2020). For freshwater species, Tudorache et al. (2008) performed an integrated research on swimming capacity and energy use in seven European freshwater fish species, showing that U_{crit} and oxygen consumption both were positively correlated to migration capacity. In tilapia, U_{crit} has been shown to be highly correlated with maximum metabolic rate (McKenzie et al., 2003). It therefore is reasonable to state that U_{crit} is strongly positively correlated to fitness. A previous study from our group (Mengistu et al., 2021) based on pedigree information, showed the existence of additive genetic variance for U_{crit} in Nile tilapia, and a favorable genetic correlation between U_{crit} and body weight, standard length, height and surface area at swimming test.

Selective breeding of farmed animals for economically important traits has high potential to increase aquaculture production. Traditional breeding strategies are to select families or individuals with excellent traits as parents to set up brood stocks for improvement, which were reported to be time-consuming and labor intensive, especially relatively slow or unstable when selecting economic traits that are determined by multiply segregating loci (Guo et al., 2022). Single nucleotide polymorphisms (SNPs) are DNA sequence variants within the genome. They can be used to map quantitative trait loci (QTLs) and genes affecting traits of interest. The analysis of QTLs for marker-assisted selection is more effective to accelerate genetic gain and reduces economic cost of bringing progeny to maturity compared to traditional breeding strategies (Ma et al., 2021). However, it is still unknown which quantitative trait loci (QTLs) and genes are associated with U_{crit} in tilapia. It is known that athletic performance is heritable in several species, such as human (Guth and Roth, 2013), horse (Schröder et al., 2011), and dog (Kim et al., 2018), and that variations in candidate genes involved in specific metabolic pathways play an important role in this trait. For instance, Ben-Zaken et al. (2017) suggested that the mutations C-1245T (rs35767) in insulin-like growth factor and Lys(K)-153Arg(R) in myostatin are strongly associated with skeletal muscle phenotypes in human, which are beneficial for endurance and short-distance running. Over the past decade, several studies have been conducted to identify genes that play an important role in determining the swimming performance in skeletal and cardiac muscles of fishes. In Atlantic salmon, Robinson et al. (2017) detected putative SNPs associated with aerobic exercise and swimming performance between wild and domesticated stocks. Analysis of those SNPs showed that they mapped to genes involved in energetic processes, coding for contractile filaments in the muscle and controlling cell proliferation. Raffini et al. (2020) identified genes involved in swimming behaviour,

physiology and oxygen intake differently expressed in the gill, by comparing divergent body shapes of two lake cichlid species. However, their contribution to genetic variation of swimming performance and their influence on growth performance in tilapia is not clear.

Hence, we decided to further explore the genomic architecture including SNPs and QTLs associated with swimming performance using the experimental data described in Mengistu et al. (2021). We also identify the effect of candidate SNPs and QTLs for Ucrit on growth performance. All fish from this study were genotyped using the Axiom® SNP array, which contains 65K SNP markers dispersed over the Nile tilapia reference genome (Peñaloza et al., 2020). The main objectives were: (1) to estimate the genetic correlations between U_{crit} with growth traits at early life stage, and with growth traits at later life stage using a genomic relationship matrix under hypoxia the same condition as used by most smallholder tilapia farmers, (2) to identify QTLs associated with swimming performance and (3) to estimate the effect of significant QTLs for U_{crit} on the growth traits under hypoxia. Overall, this knowledge will help in prioritizing SNPs and QTLs to select for better growth and healthier fish in tilapia breeding programs.

3.2 Material and methods

3.2.1 Ethics statement

Phenotypic measurements and sampling of the GIFT strain were conducted as part of the GIFT selective breeding program managed by WorldFish at the Aquaculture Extension Centre. All fish in the GIFT breeding population are managed in accordance with the Guiding Principles of the Animal Care, Welfare and Ethics Policy of WorldFish.

3.2.2 Experimental design and traits collection

Nile tilapia were part of the GIFT selective breeding program. Sixty families were produced using 31 males and 58 females. The mating ratios were designed for one male to at least two females. The successful mating were: 12 males each mated with one female (resulting in 12 full sib families), 12 males each mated with two females (resulting in 12 half sib groups equivalent to 24 full sib families), 4 males each mated with 3 females (four half sib groups equivalent to 12 full sib families) and 3 males each mated with 4 females (three half sib groups equivalent to 12 full sib families), which consist a total of 60 full sib families. Each full sib family was reared separately in a hapa (fine mesh net enclosure) set up in an earthen pond. A nursing hapa has a dimension of 1.0 × 1.0 × 1.0 m, each was stocked with 120 fry. Thirty to 35 fingerlings from each family were selected, anesthetized using clove oil and then individually

tagged using PIT (Passive Integrated Transponder) tags using intraperitoneal injection the position of the PIT tag was in abdominal cavity (peritoneum). The mean body size of fish at swim testing was 10.8 g and the size of the PIT tag is about 95 mg. The swimming test was performed three weeks after PIT tagging. From each family, 25 fish in a range from 5 to 10 cm standard length were selected for the swimming test using a ruler with a centimetre scale. Body weight (BWtest in g) and photographs were made one day before the swimming test. Standard length (SLtest) and height (Htest) at swimming test of the fish were obtained from each fish photographs using image analysis as described previously by Mengistu et al. (2020a). Surface area (SAtest) of Nile tilapia was calculated as:

$$\mathbf{SAtest} = \frac{\mathbf{1}}{\mathbf{4}}\boldsymbol{\pi} * \mathbf{SLtest} * \mathbf{Htest} \qquad (\mathbf{1})$$

Fish feeding was stopped 24 h before the beginning of the swimming test. The fish were acclimatized for one hour in the swimming flume without flow. The critical swimming test was executed using a Brett type rectangular oval shape raceway with a swimming compartment on one side and a propellor for inducing flow on the other side as described by Mengistu et al. (2021). In short, the swim flume measured 230 cm in length and 90 cm in width with a water depth of 40 cm, and was equipped with a Minn Kota Terrova 80 lbs propeller. After every 30 mins, the velocity was increased to the next speed interval until the fish fatigued. At each setting, the average water flow velocity was recorded using a FP111 Global Water Flow Probe (Mengistu et al., 2021, supplementary Table 1 for mean water velocities and standard deviations at each propeller speed setting). The mean water temperature in the tank was 28.3 ± 0.6 °C during the swim test. The swimming test takes maximally 4.5 h with 9 propeller speed level. A fish fatigue is defined when it touched the back fence and could not be stimulated to continue swimming. Each fatigued fish was taken out immediately, and fatigue time was recorded. Fatigue time was used to calculate the absolute critical swimming speed (U_{crit}) as below (Brett, 1964):

$$\mathbf{U_{crit}} = \mathbf{U_{-1}} + \left(\frac{\mathbf{t}}{\boldsymbol{\Delta}\mathbf{t}}\right)\boldsymbol{\Delta}\mathbf{U} \qquad (\mathbf{2})$$

where $\mathbf{U_{-1}}$ is the highest velocity maintained for the full-time period, $\mathbf{t}$ is the time to fatigue at final velocity level in minutes, $\boldsymbol{\Delta}\mathbf{t}$ is the time each velocity level is maintained at (= 30 min), $\boldsymbol{\Delta}\mathbf{U}$ is velocity increment in cm/s.

The caudal fin clip tissue samples were cut from each individual fish using a 3 mm diameter hole punch before stocking in the pond. Fin clips were preserved in 95% ethanol and stored at −20 °C until DNA extraction. The experimental design and data collection is published by Mengistu et al. (2021). After the swimming test, fish were

stocked in a non-aerated pond for grow-out and harvested after 145 or 146 days to keep similar hypoxic conditions as in most smallholder tilapia farming systems. The dissolved oxygen level varied from 0.91 mg/L to 6.21 mg/L measured every 2 h for 24 h using Eco-Sense® DO200A reported in our previous study (Yu et al., 2021). The stocking and harvest weight (Harw) were recorded and daily growth coefficient (DGC) was calculated as below (Bureau et al., 2000).

$$\mathbf{DGC} = \left[\frac{\sqrt[3]{\mathbf{harvest\ weight}} - \sqrt[3]{\mathbf{stocking\ weight}}}{\mathbf{time\ in\ days}}\right] \times \mathbf{100} \qquad (\mathbf{3})$$

3.2.3 SNP genotyping and quality control

A SNP array is a powerful high-throughput genotyping tool to characterize genome-wide single-nucleotide polymorphisms (SNPs). The DNA was extracted and genotyped by Identigen (Dublin, Ireland) using an Axiom® SNP array, which contains 65K SNP markers dispersed over the Nile tilapia reference genome (Peñaloza et al., 2020). The raw data from SNP array genotyping was imported to the Axiom analysis Suite version 4.0.3 software for genotype calling and quality control. Data was filtered to meet a dish quality control (DQC) >0.82 and call rate for samples (CR) >0.93, respectively. Next, a second quality control step was applied based on per SNP call rate and minor allele frequency (MAF) statistics using PLINK v1.90 (Purcell et al., 2007). A genotype call rate threshold (>90%) was set for SNP filtering. SNPs with MAF higher than 5% were retained. A total of 1388 fish and 51,438 SNPs were used for subsequent analyses.

3.2.4 Descriptive statistics and genetic parameters estimation

Basic statistics of phenotype data was analysed in R (4.0.2). The difference of traits between male and female were compared by unpaired two-samples t-test. We built the genomic relationship matrix (GRM) with 51,438 SNPs using calc_grm program with vanraden2 option (Calus and Vandenplas, 2013). Variance components and heritabilities for traits including U_{crit}, SLtest, Htest, SAtest, BWtest, Harw and DGC were estimated using univariate models based on residual maximum likelihood method using ASReml version 4.1.0 (Gilmour et al., 2015). The following model was applied:

$$\mathbf{y} = \mathbf{Xb} + \mathbf{Z_1 a} + \mathbf{Z_2 c} + \mathbf{e} \qquad (\mathbf{4})$$

where $\mathbf{y}$ is a vector with observations for one trait, being U_{crit}, SLtest, Htest, SAtest, BWtest, Harw and DGC, and $\mathbf{b}$ is a vector with fixed effects. Test day and sex are significant factors for SLtest, Htest, SAtest and BWtest; hence, only test day was fitted as class variable for U_{crit}, while age at harvest as a covariate and sex as a class variable for Harw; weight at stocking as a covariate and sex as a class variable for

DGC. **a** is a vector of the additive genetic effects of individuals and was assumed to be distributed as *N* (**0**, **G**σ_a^2), with **G** the genomic relationship matrix and σ_a^2 is the additive genetic variance; **c** is a vector of the common environmental effects, and was assumed to be distributed as N (**0**, **I**σ_c^2), **e** is a vector with the random residual and is assumed to be distributed as *N* (**0**, **I**σ_e^2), with **I** the identity matrix and σ_e^2 is the residual variance; **X**, $\mathbf{Z_1}$ and $\mathbf{Z_2}$ are design matrices assigning trait values to the fixed effects, additive genetic effects and common environmental effects. Heritability (h^2) was estimated as the ratio of additive genetic variance to the phenotypic variance. Phenotypic and genetic correlations between U_{crit} and traits such as SLtest, Htest, SAtest, BWtest, Harw and DGC were estimated based on a bivariate linear model. The log-likelihood for the bivariate model did not converge when the common environmental effect was fitted as random effect, therefore the common environmental effect was excluded in all bivariate models. The fixed effects were the same as in the univariate models for U_{crit}, SLtest, Htest, SAtest, BWtest, Harw and DGC traits.

3.2.5 Association analysis for swimming performance

As a Gaussian distribution is assumed for phenotypes in an association test of quantitative trait (Goh and Yap, 2009), the normality of traits including U_{crit}, Harw and DGC was tested with a Shapiro-wilk test. Because it is impossible to completely follow an absolute normal distribution in most cases, we normalized them by the square root method (McDonald, 2009). To test significance of factors in the experiment, a linear model was conducted using Stepwise Algorithm (Neerchal et al., 2014). The fix factors were the same as in the univariate models. Only test day was fitted as class variable for U_{crit}; age at harvest as a covariate and sex as a class variable for Harw, while stocking weight as a covariate and sex as a class variable for DGC. Once the most appropriate linear model had been fitted, residuals were extracted for the subsequent association analysis (Gondro et al., 2013).

Population structure and kinships can be confounding factors in genome-wide association studies (Hoffman, 2013). The top five principal components were added as covariates and included in subsequent GWAS models as fixed effect in this association model, since we observed a slight family structure. All SNPs that passed the quality control were used to generate the genomic relationship matrix.

Animal phenotypes and genotypes were jointly analysed to identify genomic regions associated with U_{crit}, Harw and DGC. The following model was applied:

$$\mathbf{y} = \mathbf{W\alpha} + \mathbf{Z\mu} + \mathbf{x\beta} + \boldsymbol{\varepsilon} \qquad (\mathbf{5})$$

Where **y** is the vector of each trait from (U_{crit}, Harw and DGC); **α** is a vector of associated fixed effects, **μ** is the vector of additive genetic effects, **W** and **Z** are the corresponding design matrices, **x** being the vector of SNP genotypes and **β** their associated effects, **ε** is vector with the residual effects. Analyses were conducted using the GEMMA software (Zhou and Stephens, 2012). As the Bonferroni method is overly conservative, SimpleM method which based on the effective number of independent tests (Gao et al., 2010), was used to calculate the suggestive (5.15E-05) and genome-wide significance (2.57E-06) thresholds. The empirical p-values are based on the Wald tests. Manhattan and quantile-quantile (Q-Q) plots were generated through the "qqman" package (https://cran.r-project.org/web/packages/qqman/). The inflation factor λ was calculated to indicate the influence of population structure in the association analyses.

3.2.6 Candidate genes in QTLs

Candidate regions associated with swimming performances were characterized within a 200 kb window size (100 kb upstream and downstream) flanking the candidate SNPs. LD and haplotype blocks were analysed with LDBlockShow (Dong et al., 2020), while r^2 >0.8 as cut-off for blocks. Candidate genes were defined as genes located in the same haplotype blocks as the candidate SNPs or nearby candidate SNPs if no block is present. Afterwards, candidate genes were functionally annotated based on the latest tilapia reference genome (O_niloticus_UMD_NMBU) downloaded from NCBI Genome database (https://www.ncbi.nlm.nih.gov/assembly/GCF_001858045.2). To better understand the function of tilapia genes, we performed a BLAST against zebra fish (*Danio rerio*) proteins based on the genome (GCF_000002035.6_GRCz11), using a threshold of E-value < 1e-6.

3.2.7 The effect of candidate SNPs in swimming performance on growth

In order to investigate the influence on growth of candidate SNPs associated with U_{crit}, we estimated the genetic association between SNPs and growth traits (DGC, Harw) based on generalized linear model in SNPassoc (González et al., 2007). The best fitting genetic model was evaluated based on Akaike Information Criterion (AIC) score. The significant threshold was defined as adjusted P-value (FDR) < 0.05.

3.3 Results

3.3.1 Descriptive statistics

A total of 1500 fish (60 families and 25 fish per family) were tested for swimming performance. After the swimming test, fish were stocked and allowed to grow out in non-aerated pond, except 260 fish that were dead before stocking. Comparing the U_{crit} of the dead fish to those surviving, showed no significant difference (P value = 0.29). Descriptive statistics of U_{crit}, SLtest, Htest, BWtest, SAtest, Harw and DGC for further analyses are presented in Table 1. U_{crit} was not significant different between males and females. However, all other traits presented a significantly higher mean value in males compared to that in females.

Table 1. Summary statistics of traits for all swimming tested fish

Trait	Sex	No.	Mean	SE	t-value	Effect of sex (P value)
U_{crit} (cm/s)	male	680	69.05	0.21	-0.12	NS (0.91)
	female	702	69.09	0.20		
SLtest (cm)	male	683	7.24	0.02	4.20	S (2.84E-05)
	female	705	7.11	0.02		
Htest (cm)	male	683	2.74	0.01	5.01	S (6.15E-07)
	female	705	2.67	0.01		
BWtest (g)	male	683	11.07	0.10	4.56	S (5.56E-06)
	female	705	10.45	0.09		
SAtest (cm^2)	male	683	15.67	0.10	4.82	S (1.63E-06)
	female	705	15.02	0.09		
Harw (g)	male	537	457.1	3.82	16.4	S (< 2.2e-16)
	female	589	379.5	2.78		
DGC ($g^{1/3}$/d)	male	537	3.18	0.01	9.62	S (< 2.2e-16)
	female	589	3.00	0.01		

S: significant; NS: non-significant; standard length (SLtest), height (Htest), surface area (SAtest); body weight (BWtest); harvest weight (Harw); daily growth coefficient (DGC) (P < 0.05 was set as significance threshold).

3.3.2 SNP distribution, allele frequency and family structure

In total 55,119 SNPs were exported from quality control in the Axiom Analysis Suite software (v4.03) and further investigated for their distribution, minor allele frequency and family structure. Those SNPs were distributed across the whole genome (Fig. 1a). The highest number of SNPs (4457) was seen for LG7, while the lowest number of SNPs (249) was found on an unplaced contig. The SNP distribution was mostly consistent with the physical length of the linkage groups according to the

Oreochromis niloticus genome assembly (O_niloticus_UMD_NMBU). The majority of the SNPs belonged to the common (MAF ≥ 0.3) and intermediate (0.3 > MAF ≥ 0.1) groups, which consisted of 19,757 and 21,198 SNPs respectively (Supplementary Fig. 1).

The PCA based on 51,438 informative SNPs showed genetic variation amongst those 60 GIFT families. The first, second and third components explained 15.6 %, 13.2 % and 11.5% genotype variation, respectively (Fig. 1b). However, it seems that there are several clusters that represent families.

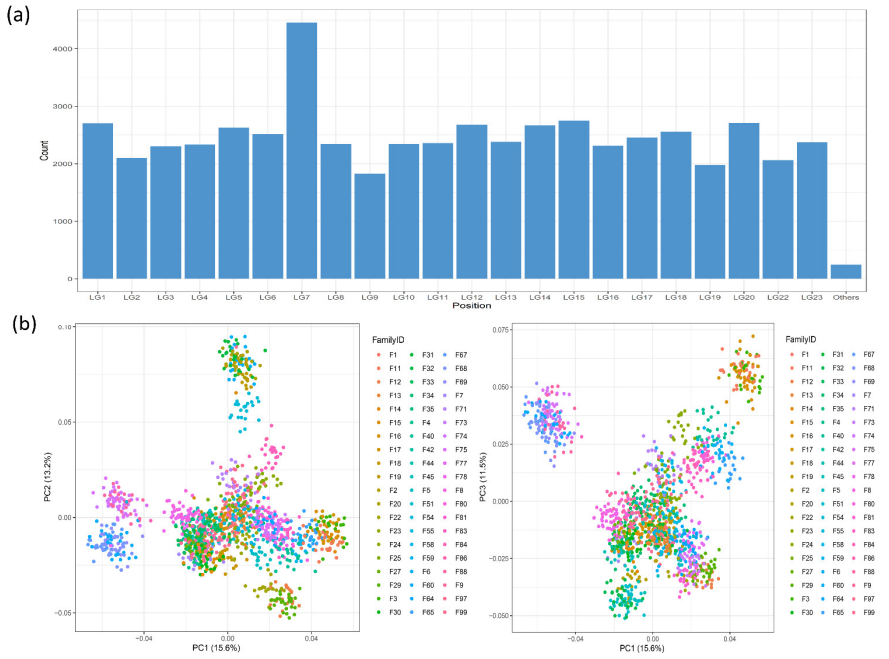


Fig. 1. SNP distribution generated from SNP array (a), PC plot on SNPs (b), each dot represents one individual.

3.3.3 Phenotypic and genetic parameter estimation with different traits at swimming test

Variance components and heritabilities from univariate models for different traits including U_{crit}, SLtest, Htest, SAtest and BWtest are presented in Table 2. The h^2 for U_{crit} was 0.31 with standard error 0.043. The variance explained by common environment was not significant ranging from 1.20E-06 to 3.10E-03. The h^2 for the remaining four traits at the swimming test (SLtest, Htest, SAtest and BWtest) were relatively close, ranging from 0.23 to 0.27. The h^2 estimates for harvest weight and

DGC were moderate with 0.29 ± 0.047 and 0.30 ± 0.046, respectively. Phenotypic and genetic correlations are shown in Table 3. The genetic correlation between U_{crit} and other traits at the time of swimming test ranged from 0.35 to 0.43, indicating that fish with better swimming capacity usually show a longer standard length, larger body height, weight and body area compared to fish with poorer swimming capacity. The estimated r_g between U_{crit} and harvest weight (−0.13 ± 0.13) and U_{crit} and DGC (−0.26 ± 0.13) were slightly negative but not different from zero, suggesting that U_{crit} might have a negative genetic correlation to body weight at harvest and growth until harvest.

Table 2. Genetic variance components and heritability for traits at swimming test in GIFT population.

Traits	σ_a^2	σ_e^2	σ_c^2	σ_p^2	h^2
U_{crit}	5.15	11.72	1.20E-06	16.88	0.31 (0.043)
SLtest	0.05	0.18	1.70E-04	0.23	0.24 (0.042)
Htest	0.01	0.03	2.40E-05	0.04	0.26 (0.042)
SAtest	1.06	3.47	3.05E-03	4.53	0.23 (0.042)
BWtest	1.23	3.38	3.10E-03	4.61	0.27 (0.043)
Harw	1515.89	3652.26	2.23	5170.40	0.29 (0.047)
DGC	0.02	0.05	0.35E-09	0.08	0.30 (0.046)

U_{crit} when only test day was fitted as class variable in the model; SLtest, Htest, SAtest and BWtest were estimated when test day and sex were included in the model as class variables; age at harvest as a covariate and sex as a class variable for Harw; weight at stocking as covariate and sex as a class variable for DGC.

Table 3. Genetic and phenotypic correlations between U_{crit} and other traits at swimming test, harvest weight, daily growth coefficient.

Traits	r_g	r_p
BWtest	0.43 ± 0.10	0.36 ± 0.03
SLtest	0.42 ± 0.11	0.35 ± 0.03
Htest	0.35 ± 0.11	0.30 ± 0.03
SAtest	0.39 ± 0.11	0.33 ± 0.03
Harw	-0.13 ± 0.13	0.06 ± 0.04
DGC	−0.26 ± 0.13	−0.01 ± 0.04

In a bivariate model, the fixed effect was the same as in the univariate models above except the common environmental effect was excluded in all models.

3.3.4 Genome-wide association study for swimming performance and growth

The linear mixed model was implemented to identify QTLs associated with critical swimming speed and growth. The P values of corrected thresholds for 5% genome-wide significant levels and suggestive association were 2.57E-06 and 5.15E-05, respectively. The genome-wide association results are shown in Fig 2, while the inflation factor lambda (λ) was estimated to be 0.97, suggesting there is little population stratification present in the association results. In total, nine SNPs located on LG12, LG13, LG14 and LG19 exceeded the genome-wide suggestive threshold for U_{crit} (Fig. 2a), while none exceeded the genome-wide significant threshold. Five out of 9 SNPs were located between 12.18 and 19.89 Mb on LG14. Two SNPs were located between 18.85 and 18.94 Mb on LG13, while the remaining two SNPs were located on LG19:3659540 and LG12:24066436 (Supplementary Table 1). Five of the suggestive SNPs showed the minor allele (ranging from 0.074 to 0.485) to be associated with improved swimming performance and indicated potential for maker-assisted selection for this trait. For harvest weight (Fig. 2b), only four SNPs on LG19 exceeded the genome-wide suggestive threshold. However, there is a peak located on LG16 between 22.55 Mb to 38.36 Mb for DGC (Fig. 2c), which include in total 7 SNPs over the suggestive threshold. Notably, there were three overlapping peaks above the suggestive line from LG7, LG16 and LG21 for both harvest weight and DGC, although a peak on LG11 did not attain any statistical significance.

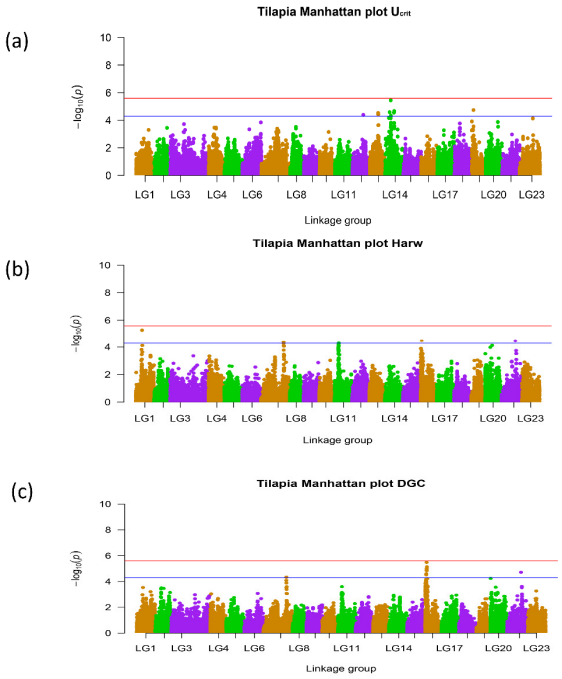


Fig. 2. Manhattan plot for Ucrit (a), Harw (b) and DGC (c). The orange and blue horizontal line represent the genome-wide significance (2.57E-06) and suggestive significance threshold value (5.15E-05) respectively. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

3.3.5 Candidate genes in the QTL regions

Candidate genes were defined as genes located in the same haplotype blocks as the candidate SNPs or nearby candidate SNPs if no block is present (as showed in Supplementary Fig. 2). A summary of all candidate genes and their functions is shown in Table 4. The SNPs with the highest significance for U_{crit} were located in an intron of the *hip1* (huntingtin interacting protein 1) gene on LG14. *Hip1* is mainly involved in regulating the central nervous system and body length. Other candidate genes on LG14 were *limk1a*, *elna*, *lsamp, aipl1* and *dner*. Candidate genes on LG13 were *pprc1* and *rrp12*, while two other candidate genes located on LG19 and LG12 were *hectd1* and *smyd1*, respectively. The functional annotation for these candidate genes suggests the involvement of several relevant biological processes, including neuronal growth, muscle activity, cardiovascular development and oxygen/hypoxia regulation.

Table 4. Candidate genes from suggestive SNPs and their biological functions.

LG	rs	pos	Candidate genes	Functional annotation
14	AX-317442766	12189511	*hip1*	central nervous system, body length
19	AX-317595114	3659540	*hectd1*	hypoxia regulator
14	AX-317032241	19888564	LOC112842187 (ncRNA)	cardiovascular development
			LOC100693040(*limk1a*)	
			LOC102081909 (ncRNA)	
			LOC100694912 (*elna*)	
14	AX-317417214	9648027	*lsamp*	neuronal growth
14	AX-317424375	19382001	*aipl1*	photoreceptor
			trarg1	
			LOC109204983 (ncRNA)	
			gosr1	
13	AX-317391485	18854572	*pprc1*	exercise-related muscle activity
13	AX-317415431	18943978	*rrp12*	oxygen adaptation
12	AX-323027008	24066436	LOC100708974 (*smyd1b*)	skeletal and cardiac muscles activity
14	AX-317418538	12496362	*dner*	neuronal and glial differentiation

Genes names in brackets were annotated based on a BALST against *Danio rerio.*

3.3.6 The effect of suggestive SNPs from U_{crit} on growth

To investigate the effect of candidate SNPs on growth, the significance threshold was defined with FDR multiple testing value < 0.05. The summary for the effect of the nine suggestive SNPs with DGC is shown in Supplementary Table 2. One significant SNP with DGC (AX-317442766) followed a dominant genetic model, which means the genotype homozygous for the major allele (TT) had significantly higher DGC (3.12) compared to the mean DGC (3.07) of CT or CC genotype. Two SNPs (AX-317595114 and AX-317418538) were significantly associated with DGC following an over-dominant genetic model. For SNP AX-317595114, the heterozygous genotype (TC) had a significantly higher DGC (3.11) compared to the mean DGC of CC and TT (3.06) genotype. However, for SNP AX-317418538, the heterozygous genotype (AG) had a significantly lower DGC (3.04) compared to the mean DGC of GG and AA (3.10). From the summary for the effect of the nine SNPs with harvest weight (Supplementary Table 3), we found that three SNPs (AX-317442766, AX-317595114 and AX-317418538) were significant, the same for DGC. We also investigated the effect size of the significant SNPs, to understand the interplay amongst U_{crit}, harvest weight and DGC. We observed that the TT genotype of SNP AX-317442766 had a lower mean U_{crit} than the genotypes CT and CC but had the highest mean harvest weight and DGC (as shown in Fig. 3a). The same is observed for AX-317418538 (Fig. 3c), in which the GG genotype had a lower mean U_{crit} than genotype AA and AG but had the highest mean harvest weight and DGC. For SNP AX-317595114 (Fig. 3b), the TT genotype had

a highest mean U_{crit}, but also a lowest harvest weight and DGC compared to genotype TC and CC. Overall, three of these 9 SNPs were significantly associated with both harvest weight and DGC, and SNP genotypes that presented a low mean U_{crit} had a high mean harvest weight and DGC.

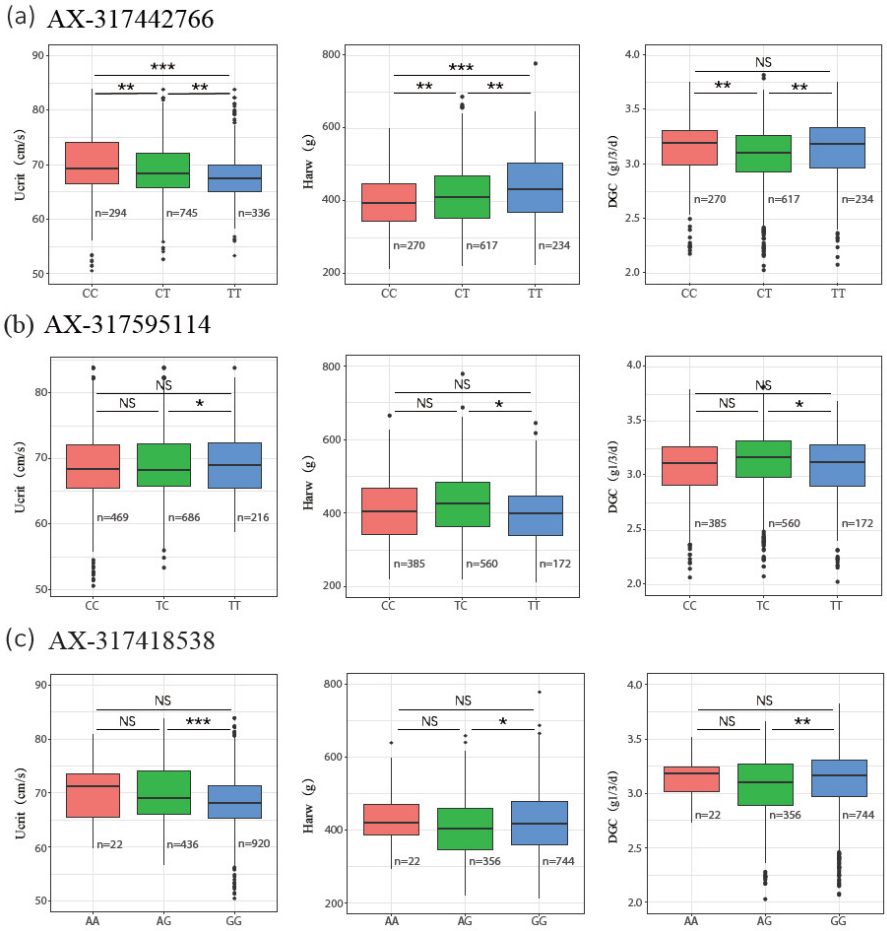


Fig. 3. The effect estimation of candidate SNPs on growth with AX-317442766 (a), AX-317418538 (b) and AX-31739145 (c).

3.4 Discussion

Following the earlier study of swimming performance in tilapia (Mengistu et al., 2021), we re-examined the genetic correlations between U_{crit} with body size traits at the swimming test and with growth traits after a growing-out period in a non-aerated pond using a genomic relationship matrix. To better understand the genomic architecture of swimming performance in Nile tilapia, we identified QTLs associated

with critical swimming speed, and estimated the effect of candidate SNPs for U_{crit} on the growth traits.

3.4.1 Heritability and genetic correlation

Heritabilities for all traits investigated in this study was moderate to high. Our results confirm those from the earlier study by Mengistu et al. (2021) and provide further evidence and support that swimming performance is heritable and potentially applicable in genomic selection. The standard error of the heritability estimates was smaller using the genomic relationship matrix method compared to a pedigree-based relationship matrix. The genetic variance estimated for U_{crit} based on genomic relationship matrix is 5.15 in this study while 6.71 in previous study estimated by pedigree relationships by Mengistu et al. (2021). This confirms the idea that variance components are more accurately estimated when using a genomic relationship matrix, which not only utilizes the relationships of an individual with all other individuals in the analysis, but also captures the Mendelian sampling variance (Veerkamp et al., 2011). Our heritability estimate of U_{crit} is within the range with that in other species e.g. 0.24 in Guppy (*Poecilia reticulata*) (Nicoletto, 1995), 0.24 in Atlantic salmon (*Salmo salar*) (Hurley and Schom, 1984), 0.41 in threespine stickleback (*Gasterosteus aculeatus*) (Garenc et al., 1998) and 0.55 in European sea bass (*Dicentrarchus labrax*) (Vandeputte et al., 2016).

Similar to the heritability, the genetic correlations based on the genomic relationship matrix were also more precise compared to those based on pedigree information. The positive phenotypic genetic correlations between U_{crit} and SLtest, Htest, SAtest, BWtest indicate that larger fish swam faster in absolute terms during the swimming test. The observed genetic correlations between U_{crit} and Harw, and U_{crit} and DGC, suggests an interaction between growth and swimming performance. A negative genetic correlation indicates that Nile tilapia with higher U_{crit} have lower growth and harvest weight in non-aerated ponds. In juvenile common carp (*Cyprinus carpio*), when acclimatized to the lower temperature, U_{crit} was negatively correlated with feeding rate and growth rate, suggesting a trade-off between growth and exercise. But this trade-off in juvenile common carp (*Cyprinus carpio*) can disappear when acclimatized to higher temperature (Pang et al., 2016). On the other hand, a negative correlation between growth and swimming performance could occur in species that mature in their first year, like tilapia. Growth rate becomes positively correlated with U_{crit} when fish mature within a few months after hatching, this pattern has been found in European sea bass (*Dicentrarchus labrax*) (Cominassi et al., 2019), Atlantic Salmon (*Salmo salar*) and Gilthead Seabream (*Sparus aurata*) (Palstra et al., 2020). In addition, selection for high growth rate can lead to the development of cardiac

abnormalities, which has been observed in several farmed species, such as broiler chicken (Olkowski, 2007) and rainbow trout (*Oncorhynchus mykiss*) (Brijs et al., 2020). The GIFT strain has been selected for harvest weight over fifteen generations so far. The negative correlations between U_{crit} and growth traits in the study, seems to suggest that a low U_{crit} could lead to cardio-respiratory health problems in the future. These results indicate that selection for better growth in combination with selection for fish with good cardio-respiratory health is feasible in a breeding program, which will aid in providing fingerlings resilient to (periodic) hypoxic culture conditions within smallholder tilapia farming.

3.4.2 Genome-wide association study

Our results show several QTLs involved in the swimming performance in Nile tilapia. It is well-known that the larger the effect of the QTL, the higher the chance of finding a significant association for a given size of study, i.e. higher power. Our analyses showed limited numbers of clear associations, and even where candidate regions were found, the effect generally seemed limited in size. However, our study was only based on 1500 samples. Increasing the number would increase the power and stronger associations would be expected to emerge. While some medium to strong associations were found, the results also indicated that U_{crit} is a complex trait regulated by many genes with small effect.

Overall, 16 genes in high linkage disequilibrium (LD) with the identified SNPs, were found. The 65 K SNP array used in this study covers < 2% of all SNPs presented in the GIFT population (Cádiz et al., 2020). It is more likely candidate SNPs are in strong LD with the casual SNPs rather than causal SNPs. The SNP with the highest significance (LG14:12189511) was linked to the gene *hip1*, which codes for huntingtin-interacting protein. The gene has been suggested to confer narrow body shape in zebrafish (Komoike et al., 2010), whereas *HIP1* in human is presumed to affect cognitive and central nervous function (Metzler et al., 2003; Ramocki et al., 2010). Two other genes (*lsamp* and *dner*) are also potentially involved in neuronal growth and differentiation (Pimenta et al., 1995; Hsieh et al., 2013). The candidate gene *hectd1* codes for HECT domain E3 ubiquitin ligase 1. A zebrafish (*Danio rerio*) orthologue of the human and mouse *HECTD1* gene is annotated in the ZFIN database. The expression level of *HECTD1* decreased under hypoxia in human cells (Erler, 2014; Wang et al., 2020). The response to oxygen levels in human cells suggests *hectd1* in tilapia may be an interesting candidate gene under similar challenging circumstances. Similarly, the *rrp12* gene has orthologues found in both human and mouse, which have been associated with adaptation to low oxygen. The human orthologue of *RRP12* was found to be associated with living at high altitude in Asian and Amerindian

populations (Foll et al., 2014). Further candidate genes identified in the present study include *elna*, for which variant (elnasa12235 c.264T>A, p.Tyr88*) has been shown to reduced blood flow and to induce heart function abnormalities in zebrafish (*Danio rerio*) (Zorrilla, 2018). Furthermore, zebrafish (*Danio rerio*) and human orthologues of two genes (*smyd1b* and *pprc1*) are potentially involved in cardiac and exercise-related muscle activity (Li et al., 2013; Uguccioni et al., 2010).

Previous studies have shown that skeletal and cardiac muscle tissues are significantly influenced by swimming-induced exercise in several species, such as Atlantic salmon (*Salmo salar*) (Castro et al., 2013) and zebrafish (*Danio rerio*) (Rovira, 2017). Sustained swimming could increase transcriptional activity in white muscle during growth and development in rainbow trout (*Oncorhynchus mykiss*) (Palstra et al., 2013). We further confirmed that a few candidate genes were also involved in skeletal and cardiac muscle in Nile tilapia.

In summary, while the majority of genes involved in the trait probably cannot be fully identified by a study of the scope and size of the current one, we were able to identify a number of good candidates. It was found that pathways including central nervous system and neuron development, oxygen adaptation and hypoxia regulation, cardiac and exercise-related muscle activity are potentially involved in regulation of swimming performance. Overall, these 9 SNPs involved in muscle activity, cardiovascular development and angiogenesis, and oxygen/hypoxia regulation pathways, should be prioritized for marker-assisted selection in breeding program.

3.4.3 Pleiotropic SNPs with swimming performance and growth

We also investigated the impact of candidate SNPs associated with U_{crit} on growth. As the function of genes becomes better known, it is often revealed that they are involved in multiple pathways, and therefore have multiple effects on the phenotype (pleiotropy). Pleiotropy is widespread in animal breeding. When one trait is under selection, the mean of other traits also changes over generations. This response to selection could be reflected by the genetic correlation between traits (Gratten and Visscher, 2016). Knowing the genetic basis of traits and their pleiotropic effects through candidate gene and molecular pathway analysis can reveal how traits may be related at the functional genomic level.

Our results show only three of the SNPs are associated with harvest weight and DGC. These three SNPs are linked to three genes (*hip1*, *hectd1*, *dner)*. Gene *hip1* regulates body length, therefore it is not surprising that *hip1* is also associated with growth.

The gene *hectd1* is assumed to be involved in adaptation to hypoxia (Wang et al., 2020) and our previous studies have shown that hypoxia can suppress growth in Nile tilapia (Mengistu et al., 2020a; Yu et al., 2021). The candidate genes involved in cardiovascular development and exercise-related muscle activity suggested no relationship with growth. Nevertheless, they potentially can be applied as biomarkers to select fish with good cardio-respiratory health and good growth. Genes involved in neuronal growth and differentiation pathway can influence growth while others do not. After all, growth traits are also polygenic with potential involvement of thousands of QTLs (Zhang et al., 2021).

3.5 Conclusions

Moderate heritability of swimming performance was found based on genomic relationship matrix estimation. Large fish generally swim faster than smaller fish, whereas fish with better swimming performance showed slower growth later in life. Nine suggestive SNPs between genotype and swimming performance were identified. The identified QTLs indicate that swimming performance is a complex trait with pleiotropic effects on growth. Our results reveal a clear pleiotropic effect of some SNPs associated with swimming performance on growth traits including harvest weight and DGC, while other SNPs had only effect on swimming performance, but not on growth. Using these SNPs in selection has the potential to select fish with good cardio-respiratory health and good growth. Overall, our analyses shed a first light on the genomic mechanisms of swimming performance and growth in tilapia.

3.6 Funding

This work was funded by Koepon Foundation, International Fund for Agricultural Development (IFAD) Grant number: 2000001001, the CGIAR Research Program on Fish Agri-Food Systems (FISH) led by WorldFish. The FISH program is supported by contributors to the CGIAR Trust Fund. The authors highly thank the WorldFish breeding team and the Department of Fisheries Malaysia for their technical support during the experimental phase. The authors also thank to the financial support from China Scholarship Council (CSC, File No.201803250069), which has sponsored Xiaofei Yu's PhD study at Wageningen University and Research. The funding bodies played no role in design of the study and data collection, analysis, and result interpretation in the manuscript.

3.7 Authors' contributions

X.Y: Writing – original draft, Methodology, Formal analysis, Investigation, Visualization. S.M: Writing – review & editing, Conceptualization, data curation,

Formal analysis. H.M: Writing – review & editing, Conceptualization, Methodology, Investigation. A.P: Writing – review & editing, Conceptualization, Methodology, Investigation. J.B: Writing – review & editing, Conceptualization, Resources, Funding acquisition. T.T: Writing - Review & Editing, Data Curation, Methodology, Investigation. M.A.M.G: Writing – review & editing, Supervision, Investigation, Resources. H.K: Writing – review & editing, Supervision, Conceptualization, Investigation, Funding acquisition. H.J. M: Writing – original draft, Supervision, Methodology, Investigation. HAM, HK and MAMG conceived and designed the experiments. SBM and HAM contributed to animal experiment and collected biological samples, and they also supported data analysis and result interpretation. XFY, HJM wrote the manuscript. XFY, JWMB and HJM contributed to data analysis. XFY, HJM, MAMG and HK contributed to result interpretation. SBM, JWMB, HAM, JAHB, MAMG and HK contributed to manuscript revision. All authors reviewed and approved the final manuscript.

3.8 Data availability

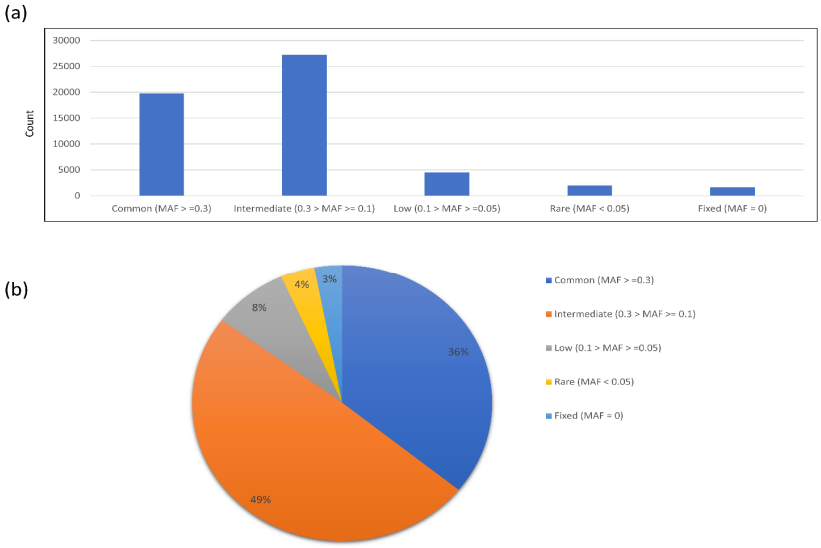
The genotype and phenotype data for swim-tested fish in this study are stored in the Harvard Dataverse repository (https://dataverse.harvard.edu/) with accession number JHQHS1. The authors declare that all data supporting the findings are available within this article and its supplementary files.

3.9 Declaration of Competing Interest

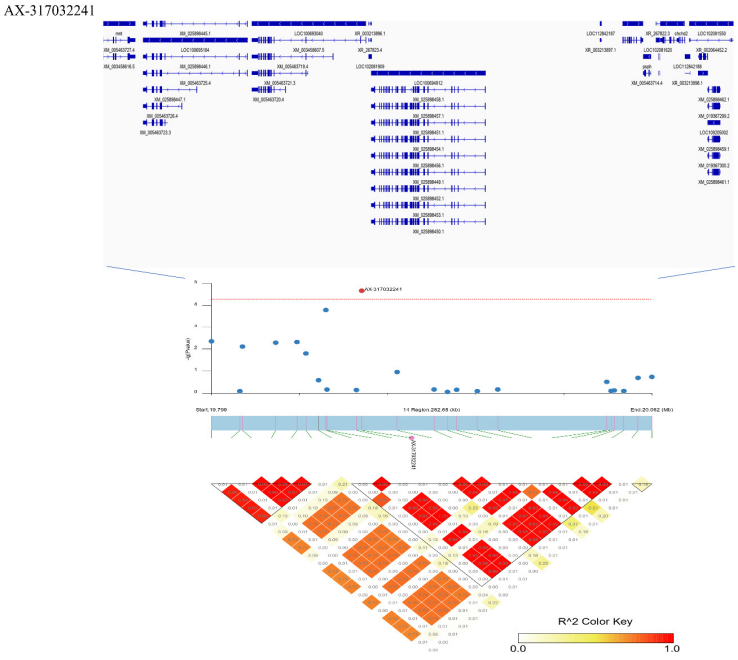
The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

3.10 Supplementary Information

For a compact layout, I have not included all Supplementary material in the thesis, only those which may assist the reader. For more information, the complete Supplementary figures and tables of the article are available at https://doi.org/10.1016/j.aquaculture.2022.738522.



Supplementary Fig. 1. Distribution of minor allele frequency for all SNPs.



Supplementary Fig. 2. Definition of candidate regions with an example SNP (AX-317032241).

Supplementary Table 1. Summary statistics of SNPs above the genome-wide suggestive level for U_{crit}.
Supplementary Table 2. Association summary between candidate SNPs and DGC with the Best-fitting model.
Supplementary Table 3. Association summary between candidate SNPs and harvest weight with the Best-fitting model.

References

Ben-Zaken, S., Meckel, Y., Nemet, D., Eliakim, A., 2017. The combined frequency of IGF and myostatin polymorphism among track & field athletes and swimmers. Growth Hormone and IGF Research 32, 29–32. https://doi.org/10.1016/j.ghir.2016.12.002

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., 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.

Brett, J.R., 1964. The Respiratory Metabolism and Swimming Performance of Young Sockeye Salmon. Journal of the Fisheries Research Board of Canada 21, 1183–1226. https://doi.org/10.1139/f64-103

Brijs, J., Hjelmstedt, P., Berg, C., Johansen, I.B., Sundh, H., Roques, J.A.C., Ekström, A., Sandblom, E., Sundell, K., Olsson, C., Axelsson, M., Gräns, A., 2020. Prevalence and severity of cardiac abnormalities and arteriosclerosis in farmed rainbow trout (Oncorhynchus mykiss). Aquaculture 526, 735417. https://doi.org/10.1016/j.aquaculture.2020.735417

Bureau, D.P., Azevedo, P.A., Tapia-Salazar, M., Cuzon, G., 2000. Pattern and cost of growth and nutrient deposition in fish and shrimp: Potential implications and applications. Avances en nutricion acuicola.

Cádiz, M.I., López, M.E., Díaz-Domínguez, D., Cáceres, G., Yoshida, G.M., Gomez-Uchida, D., Yáñez, J.M., 2020. Whole genome re-sequencing reveals recent signatures of selection in three strains of farmed Nile tilapia (Oreochromis niloticus). Scientific reports 10, 1–14.

Calus, M.P.L., Vandenplas, J., 2013. Calc_grm—a program to compute pedigree, genomic, and combined relationship matrices. Animal Breeding and Genomics Centre, Wageningen UR Livestock Research.

Cano-Barbacil, C., Radinger, J., Argudo, M., Rubio-Gracia, F., Vila-Gispert, A., García-Berthou, E., 2020. Key factors explaining critical swimming speed in freshwater fish: a review and statistical analysis for Iberian species. Scientific Reports 10, 1–12. https://doi.org/10.1038/s41598-020-75974-x

Castro, V., Grisdale-Helland, B., Helland, S.J., Torgersen, J., Kristensen, T., Claireaux, G., Farrell, A.P., Takle, H., 2013. Cardiac molecular-acclimation mechanisms in response to swimming-induced exercise in Atlantic salmon. PLoS One 8, e55056.

Cominassi, L., Moyano, M., Claireaux, G., Howald, S., Mark, F.C., Zambonino-Infante, J.L., Le Bayon, N., Peck, M.A., 2019. Combined effects of ocean acidification and temperature on larval and juvenile growth, development and swimming performance of European sea bass (Dicentrarchus labrax). PLoS ONE 14, 1–22. https://doi.org/10.1371/journal.pone.0221283

Dong, S.-S., He, W.-M., Ji, J.-J., Zhang, C., Guo, Y., Yang, T.-L., 2020. LDBlockShow: a fast and convenient tool for visualizing linkage disequilibrium and haplotype blocks based on variant call format files. Briefings in Bioinformatics 00, 1–6. https://doi.org/10.1093/bib/bbaa227

El-Sayed, A.-F.M., 2019. Tilapia culture. Academic Press.

Erler, J.T., 2014. Identification of Hypoxia-Regulated Proteins Using MALDI-Mass Spectrometry Imaging Combined with Quantitative Proteomics.

Foll, M., Gaggiotti, O.E., Daub, J.T., Vatsiou, A., Excoffier, L., 2014. Widespread signals of convergent adaptation to high altitude in Asia and America. American Journal of Human Genetics 95, 394–407. https://doi.org/10.1016/j.ajhg.2014.09.002

Gao, X., Becker, L.C., Becker, D.M., Starmer, J.D., Province, M.A., 2010. Avoiding the high Bonferroni penalty in genome-wide association studies. Genetic Epidemiology: The Official Publication of the International Genetic Epidemiology Society 34, 100–105.

Garenc, C., Silversides, F.G., Guderley, H., 1998. Burst swimming and its enzymatic correlates in the threespine stickleback (Gasterosteus aculeatus): full-sib heritabilities. Canadian Journal of Zoology 76, 680–688.

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.

Goh, L., Yap, V.B., 2009. Effects of normalization on quantitative traits in association test. BMC Bioinformatics 10. https://doi.org/10.1186/1471-2105-10-415

Gondro, C., Van Der Werf, J.H., Hayes, B., 2013. Genome-wide association studies and genomic prediction. Humana Press.

González, J.R., Armengol, L., Solé, X., Guinó, E., Mercader, J.M., Estivill, X., Moreno, V., 2007. SNPassoc: An R package to perform whole genome association studies. Bioinformatics 23, 644–645. https://doi.org/10.1093/bioinformatics/btm025

Gratten, J., Visscher, P.M., 2016. Genetic pleiotropy in complex traits and diseases: Implications for genomic medicine. Genome Medicine 8, 8–10. https://doi.org/10.1186/s13073-016-0332-x

Guo, J., Wang, A., Mao, S., Xu, X., Li, J., Shen, Y., 2022. Construction of high-density genetic linkage map and QTL mapping for growth performance in black carp (Mylopharyngodon piceus). Aquaculture 549, 737799. https://doi.org/10.1016/j.aquaculture.2021.737799

Guth, L.M., Roth, S.M., 2013. Genetic influence on athletic performance. Current opinion in pediatrics 25, 653.

Hoffman, G.E., 2013. Correcting for Population Structure and Kinship Using the Linear Mixed Model: Theory and Extensions. PLoS ONE 8. https://doi.org/10.1371/journal.pone.0075707

Hsieh, F.Y., Ma, T.L., Shih, H.Y., Lin, S.J., Huang, C.W., Wang, H.Y., Cheng, Y.C., 2013. Dner inhibits neural progenitor proliferation and induces neuronal and glial differentiation in zebrafish. Developmental Biology 375, 1–12. https://doi.org/10.1016/j.ydbio.2013.01.007

Hurley, S.M., Schom, C.B., 1984. Genetic control of swimming stamina in Atlantic salmon (Salmo salar). Canadian journal of genetics and cytology 26, 57–61.

Kieffer, James D., 2010. Perspective - Exercise in fish: 50+ years and going strong. Comparative Biochemistry and Physiology - A Molecular and Integrative Physiology 156, 163–168. https://doi.org/10.1016/j.cbpa.2010.02.009

Kim, J., Williams, F.J., Dreger, D.L., Plassais, J., Davis, B.W., Parker, H.G., Ostrander, E.A., 2018. Genetic selection of athletic success in sport-hunting dogs. Proceedings of the National Academy of Sciences of the United States of America 115, E7212–E7221. https://doi.org/10.1073/pnas.1800455115

Komoike, Y., Fujii, K., Nishimura, A., Hiraki, Y., Hayashidani, M., Shimojima, K., Nishizawa, T., Higashi, K., Yasukawa, K., Saitsu, H., Miyake, N., Mizuguchi, T., Matsumoto, N., Osawa, M., Kohno, Y., Higashinakagawa, T., Yamamoto, T., 2010. Zebrafish gene knockdowns imply roles for human YWHAG in infantile spasms and cardiomegaly. Genesis 48, 233–243. https://doi.org/10.1002/dvg.20607

Li, H., Zhong, Y., Wang, Z., Gao, J., Xu, J., Chu, W., Zhang, J., Fang, S., Du, S.J., 2013. Smyd1b is required for skeletal and cardiac muscle function in zebrafish. Molecular biology of the cell 24, 3511–3521.

Ma, A., Huang, Z., Wang, X. an, Xu, Y., Guo, X., 2021. Identification of quantitative trait loci associated with upper temperature tolerance in turbot, Scophthalmus maximus. Scientific Reports 11, 1–12. https://doi.org/10.1038/s41598-021-01062-3

McDonald, J.H., 2009. Handbook of biological statistics. sparky house publishing Baltimore, MD.

McKenzie, D.J., Martinez, R., Morales, A., Acosta, J., Morales, R., Taylor, E.W., Steffensen, J.F., Estrada, M.P., 2003. Effects of growth hormone transgenesis on metabolic rate, exercise performance and hypoxia tolerance in tilapia hybrids. Journal of Fish Biology 63, 398–409. https://doi.org/10.1046/j.1095-8649.2003.00162.x

Mengistu, Samuel B., Mulder, H.A., Benzie, J.A.H., Khaw, H.L., Megens, H.J., Trinh, T.Q., Komen, H., 2020a. 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

Mengistu, Samuel Bekele, Mulder, H.A., Benzie, J.A.H., Komen, H., 2020b. A systematic literature review of the major factors causing yield gap by affecting growth, feed conversion ratio and survival in Nile tilapia (Oreochromis niloticus). Reviews in Aquaculture 12, 524–541. https://doi.org/10.1111/raq.12331

Mengistu, S.B., Palstra, A.P., Mulder, H.A., Benzie, J.A.H., Trinh, T.Q., Roozeboom, C., Komen, H., 2021. Heritable variation in swimming performance in Nile tilapia (Oreochromis niloticus) and negative genetic correlations with growth and harvest weight. Scientific Reports 11, 1–11. https://doi.org/10.1038/s41598-021-90418-w

Metzler, M., Li, B., Gan, L., Georgiou, J., Gutekunst, C.A., Wang, Y.T., Torre, E., Devon, R.S., Oh, R., Legendre-Guillemin, V., Rich, M., Alvarez, C., Gertsenstein, M., McPhersons, P.S., Nagy, A., Wang, Y.T., Roder, J.C., Raymond, L.A., Hayden, M.R., 2003. Disruption of the endocytic protein HIP1 results in neurological deficits and decreased AMPA receptor trafficking. EMBO Journal 22, 3254–3266. https://doi.org/10.1093/emboj/cdg334

Neerchal, N.K., Morel, J.G., Huang, X., Moluh, A., 2014. A stepwise algorithm for generalized linear mixed models, in: SAS Global Forum. pp. 1822–2014.

Nicoletto, P.F., 1995. Offspring quality and female choice in the guppy, Poecilia reticulata. Animal Behaviour 49, 377–387.

Olkowski, A.A., 2007. Pathophysiology of heart failure in broiler chickens: Structural, biochemical, and molecular characteristics. Poultry Science 86, 999–1005. https://doi.org/10.1093/ps/86.5.999

Palstra, A.P., Beltran, S., Burgerhout, E., Brittijn, S.A., Magnoni, L.J., Henkel, C. V, Jansen, H.J., Van Den Thillart, G.E., Spaink, H.P., Planas, J. V, 2013. Deep RNA sequencing of the skeletal muscle transcriptome in swimming fish. PloS one 8, e53171.

Palstra, A.P., Kals, J., Böhm, T., Bastiaansen, J.W.M., Komen, H., 2020. Swimming Performance and Oxygen Consumption as Non-lethal Indicators of Production Traits in Atlantic Salmon and Gilthead Seabream. Frontiers in Physiology 11, 1–14. https://doi.org/10.3389/fphys.2020.00759

Palstra, A.P., Planas, J. V., 2012. Swimming Physiology of Fish: Towards Using Exercise to Farm a Fit Fish in Sustainable Aquaculture. Springer Science & Business Media. https://doi.org/10.1007/978-3-642-31049-2

Pang, X., Fu, S.J., Zhang, Y.G., 2016. Acclimation temperature alters the relationship between growth and swimming performance among juvenile common carp (Cyprinus carpio).

Comparative Biochemistry and Physiology -Part A : Molecular and Integrative Physiology 199, 111–119. https://doi.org/10.1016/j.cbpa.2016.06.011

Peñaloza, C., Robledo, D., Barría, A., Trinh, T.Q., Mahmuddin, M., Wiener, P., Benzie, J.A.H., Houston, R.D., 2020. Development and validation of an open access SNP array for Nile Tilapia (oreochromis niloticus). G3: Genes, Genomes, Genetics 10, 2777–2785. https://doi.org/10.1534/g3.120.401343

Pimenta, A.F., Zhukareva, V., Barbe, M.F., Reinoso, B.S., Grimley, C., Henzel, W., Fischer, I., Levitt, P., 1995. The limbic system-associated membrane protein is an Ig superfamily member that mediates selective neuronal growth and axon targeting. Neuron 15, 287–297. https://doi.org/10.1016/0896-6273(95)90034-9

Plaut, I., 2001. Critical swimming speed: its ecological relevance. Comparative Biochemistry and Physiology Part A: Molecular & Integrative Physiology 131, 41–50.

Ponzoni, R.W., Nguyen, N.H., Khaw, H.L., Hamzah, A., Bakar, K.R.A., Yee, H.Y., 2011. Genetic improvement of Nile tilapia (Oreochromis niloticus) with special reference to the work conducted by the World Fish Center with the GIFT strain. Reviews in Aquaculture 3, 27–41. https://doi.org/10.1111/j.1753-5131.2010.01041.x

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A.R., Bender, D., Maller, J., Sklar, P., De Bakker, P.I.W., Daly, M.J., 2007. PLINK: a tool set for whole-genome association and population-based linkage analyses. The American journal of human genetics 81, 559–575.

Raffini, F., Schneider, R.F., Franchini, P., Kautt, A.F., Meyer, A., 2020. Diving into divergence: Differentiation in swimming performances, physiology and gene expression between locally-adapted sympatric cichlid fishes. Molecular Ecology 29, 1219–1234. https://doi.org/10.1111/mec.15304

Ramocki, M.B., Bartnik, M., Szafranski, P., Kołodziejska, K.E., Xia, Z., Bravo, J., Miller, G.S., Rodriguez, D.L., Williams, C.A., Bader, P.I., Szczepanik, E., Mazurczak, T.T., Antczak-Marach, D., Coldwell, J.G., Akman, C.I., McAlmon, K., Cohen, M.P., McGrath, J., Roeder, E., Mueller, J., Kang, S.H.L., Bacino, C.A., Patel, A., Bocian, E., Shaw, C.A., Cheung, S.W., Mazurczak, T.T., Stankiewicz, P., 2010. Recurrent distal 7q11.23 deletion including HIP1 and YWHAG identified in patients with intellectual disabilities, epilepsy, and neurobehavioral problems. American Journal of Human Genetics 87, 857–865. https://doi.org/10.1016/j.ajhg.2010.10.019

Robinson, N.A., Timmerhaus, G., Baranski, M., Andersen, Ø., Takle, H., Krasnov, A., 2017. Training the salmon's genes: Influence of aerobic exercise, swimming performance and selection on gene expression in Atlantic salmon. BMC Genomics 18, 1–19. https://doi.org/10.1186/s12864-017-4361-7

Rovira, M., 2017. Skeletal Muscle and Cardiac Adaptations to Swimming-induced Exercise in Adult Zebrafish. PhD thesis.

Schröder, W., Klostermann, A., Distl, O., 2011. Candidate genes for physical performance in the horse. Veterinary Journal 190, 39–48. https://doi.org/10.1016/j.tvjl.2010.09.029

Tayamen, M.M., 2004. Nationwide dissemination of GET-EXCEL tilapia in the Philippines, in: Proceedings of the Sixth International Symposium On Tilapia In Aquaculture.

Thodesen, J., Rye, M., Wang, Y.-X., Yang, K.-S., Bentsen, H.B., Gjedrem, T., 2011. Genetic improvement of tilapias in China: genetic parameters and selection responses in growth of Nile tilapia (Oreochromis niloticus) after six generations of multi-trait selection for growth and fillet yield. Aquaculture 322, 51–64.

Tudorache, C., Viaene, P., Blust, R., Vereecken, H., De Boeck, G., 2008. A comparison of swimming capacity and energy use in seven European freshwater fish species. Ecology of freshwater fish 17, 284–291.

Uguccioni, G., D'souza, D., Hood, D.A., 2010. Regulation of PPAR coactivator-1 function and expression in muscle: effect of exercise. PPAR research 2010.

Vandeputte, M., Porte, J.D., Auperin, B., Dupont-Nivet, M., Vergnet, A., Valotaire, C., Claireaux, G., Prunet, P., Chatain, B., 2016. Quantitative genetic variation for post-stress cortisol and swimming performance in growth-selected and control populations of European sea bass (Dicentrarchus labrax). Aquaculture 455, 1–7. https://doi.org/10.1016/j.aquaculture.2016.01.003

Veerkamp, R.F., Mulder, H.A., Thompson, R., Calus, M.P.L., 2011. Genomic and pedigree-based genetic parameters for scarcely recorded traits when some animals are genotyped. Journal of Dairy Science 94, 4189–4197. https://doi.org/10.3168/jds.2011-4223

Wang, X., De Geyter, C., Jia, Z., Peng, Y., Zhang, H., 2020. HECTD1 regulates the expression of SNAIL: Implications for epithelial–mesenchymal transition. International Journal of Oncology 56, 1186–1198. https://doi.org/10.3892/ijo.2020.5002

Yu, X., Megens, H.J., Mengistu, S.B., Bastiaansen, J.W.M., Mulder, H.A., Benzie, J.A.H., Groenen, M.A.M., Komen, H., 2021. Genome-wide association analysis of adaptation to oxygen stress in Nile tilapia (Oreochromis niloticus). BMC Genomics 22, 1–13. https://doi.org/10.1186/s12864-021-07486-5

Zhang, Y., Gilbert, M.J.H., Farrell, A.P., 2019. Finding the peak of dynamic oxygen uptake during fatiguing exercise in fish. Journal of Experimental Biology 222. https://doi.org/10.1242/jeb.196568

Zhang, Y., Wang, Yuzhe, Li, Y., Wu, J., Wang, X., Bian, C., Tian, Y., Sun, G., Han, R., Liu, X., Jiang, R., Wang, Yanbin, Li, G., Li, W., Hu, X., Kang, X., 2021. Genome-wide association study reveals the genetic determinism of growth traits in a Gushi-Anka F2 chicken population. Heredity 126, 293–307. https://doi.org/10.1038/s41437-020-00365-x

Zhou, X., Stephens, M., 2012. Genome-wide efficient mixed-model analysis for association studies. Nature genetics 44, 821–824.

Zorrilla, M., 2018. Contribution of Elastin to Cardiovascular Development in Zebrafish.

4

Genomic analysis of a Nile tilapia strain selected for salinity tolerance shows signatures of selection and hybridization with blue tilapia (*Oreochromis aureus*)

Xiaofei Yu[1], Priadi Setyawan[1,2,3], John W.M. Bastiaansen[1], Langqing Liu[1], Imron Imron[2,3], Martien A.M. Groenen[1], Hans Komen[1] and Hendrik-Jan Megens[1]

[1]Animal Breeding and Genomics, Wageningen University & Research, the Netherlands.
[2]Research Institute for Fish Breeding, Ministry of Marine Affairs and Fisheries, Indonesia.
[3]National Research and Innovation Agency, Indonesia

Aquaculture 560 (2022): 738527

Abstract

Tilapia is a group of originally freshwater species, some of which can be tolerate a wide range of salinities and can be cultured in estuaries or brackish water ponds in polyculture with shrimp. Although the physiological processes that underly osmoregulation have been studied extensively, it is less clear how artificial selection produce adaptation to salinity stress. Here we studied the genomic architecture of an Indonesian saline-tolerant strain, called "Sukamandi", which was selected for rapid growth in brackish water. We also investigated the impact of selection for salinity tolerance on the genome. Because the Sukamandi strain was potentially derived from hybridization between Nile tilapia (*Orechromis niloticus*) and blue tilapia *(Orechromis aureus)*, we also searched for introgression signatures to understand their influence on salinity tolerance. Our results show that overall the Sukamandi strain is genetically much closer to Nile tilapia than to blue tilapia. Thirty-three salinity tolerance genes identified by *Fst*, enriched in ion transmembrane transport processes, such as MAPK3 activity, potassium ion homeostasis, ATPase activity and response to calcium ion. Comparing signatures of selection and introgression revealed that eight salinity tolerance genes, including *caprin1a*, *nucb2a*, *abcb10*, *slc12a10.1*, *cacna1ab*, *ulk2*, *slc25a24* and *cdh1* were strongly selected (top 1% signal windows) based on genome-wide scans, while five (*slc12a10.1, zgc:153039, slc9a2, slc25a24, cdh1)* out of thirty-three genes that have been introgressed from blue tilapia and selected (above top 5% signal windows) in the Sukamandi strain. Our findings not only contribute to understanding the evolution of salinity tolerance in fish, but, more generally, provide an interesting model for hybrid introgression in a farmed fish species.

Keywords: Tilapia, adaptation, salinity tolerance, artificial selection, inter-species introgression

4.1 Introduction

Farmed tilapia, including Nile tilapia (*Oreochromis niloticus*), and some other cichlid species, are currently the most important fish species in aquaculture in the tropics and subtropics (Mjoun et al., 2010a; Prabu et al., 2019). Limited freshwater resources and competition with agriculture and other urban activities, has increased the pressure to develop aquaculture in brackish water and sea water in many countries (El-Sayed, 2006). Tilapia is an excellent candidate for aquaculture in brackish or saline environments due to their ability to tolerate wide range of environmental conditions (Mjoun et al., 2010b; El-Sayed, 2019). Saline-tolerant tilapia is also well-suited for polyculture with saltwater shrimp, which can have both environmental and economic benefits by improving feed efficiency and reducing the effects of diseases (Junior et al., 2012; Martínez-Porchas et al., 2010; Fitzsimmons and Shahkar, 2017). Within the genus *Oreochromis* there are a number of saline-tolerant species, such as Mozambique tilapia (*Oreochromis mossambicus*) and blue tilapia (*Oreochromis aureus*) (El-Sayed, 2006). In contrast to Nile tilapia, these tilapia species have been a limited history of selection for high yield (Jaspe and Caipang, 2011; Zak et al., 2014). However, Nile tilapia, which has a long history of selection for production traits does not tolerate high salinity well (Lawson and Anetekhai, 2011; Ponzoni et al., 2011).

To address this limitation, a saline-tolerant and highly productive tilapia has been developed by the aquaculture research institute, Research Institute for Fish Breeding (RIFB) in Indonesia and named “Sukamandi”. Starting with ~100 fish first spawned for multiplication in early 2008, RIFB conducted a breeding program for the Sukamandi strain based on own performance for harvest weight after 120 days of growth in brackish water ponds. The estimated cumulative response after four generations of selection for harvest weight is 39.6 g and 38.8 g for males and females, respectively. The Sukamandi strain also shows stable growth under high and variable salinity (30 to 58 ppt, unpublished results). In Indonesia, variation in salinity is driven by heavy annual cycles of rain and drought. Setyawan and Robisalmi (2014) showed that this strain has a survival rate of 82.7%, much higher than the 41.07% recorded for the Nile tilapia, during a three months grow-out period in brackish water ponds. However, the harvest weights were not significantly different. The history of the Sukamandi strain is not fully documented, but it is thought to be derived from blue tilapia (*Oreochromis aureus*). A Nile tilapia component of the Sukamandi strain has been provided by RIFB as likely. Due to incomplete documentation of the breeding history, it is unknown to have been a deliberate choice or not.

Salinity is an important environmental factor for aquatic organisms, and many fish species are well-known for their remarkable abilities to tolerate a range of salinities. Physiological responses associated with adaptation to high salinity include water and ion transport in the major osmoregulatory tissues (i.e. gill, kidney and intestine) (Laverty and Skadhauge, 2012). Chloride cells (also known as mitochondrion-rich cell) in the gill epithelium are important osmoregulatory sites in teleost species, with sodium-potassium ATPase (Na+/K+-ATPase, or NKA) and Na+/K+/2Cl−co-transporter (NKCC) as the main ion transporting proteins (Cnaani et al., 2011). Different isoforms of the Na+/K+- ATPase protein show differences in abundance in the gill chloride cells between freshwater and seawater (McCormick et al., 2009). Na+/K+-ATPase is regulated by growth hormone and prolactin, which play important roles in salinity adaptation (Kalujnaia et al., 2007).

Understanding adaptation to salinity stress is required to optimize selective breeding of tilapia, especially when the environment is variable, and a rapid adaptive response is required. Adaptation can be identified in those genes that are subject to positive Darwinian selection (Shapiro and Alm, 2008). Genome variations in aquacultural species, especially those which are under artificial selection, are particularly interesting as they may be correlated with economic traits, such as growth (Sciara et al., 2018), meat quality (Ali et al., 2020) and disease resistance (Robledo et al., 2018). So far, several studies have investigated genomic variations, mostly single nucleotide polymorphisms (SNPs), that contribute to salinity tolerance in tilapia (Rengmark et al., 2007; Gu et al., 2018a, 2018b; Jiang et al., 2019). Salinity tolerance has also been studied in several other fish species, including rainbow trout (*Oncorhynchus mykiss*) (Le Bras et al., 2011), Atlantic salmon (*Salmo salar*) (Norman et al., 2012) and stickleback (Kusakabe et al., 2017). In addition, a series of genes and proteins including prolactin (*PRL*) (Breves et al., 2013), growth hormone (*GH1*) (Deane and Woo, 2009), insulin-like growth factor1 (*IGF1*) (Yan et al., 2020), and plasma-membrane Ca2+-ATPases (PMCAs) (Rengmark et al., 2007) were found to confer tolerance to high salinity. Identification of variation in the genomic regions and candidate genes that confer salinity tolerance will give a better understanding about the genetic basis of salinity tolerance in fish (Kusakabe, 2017), and may subsequently guide selective breeding.

Genetic variation can cross species boundaries through hybridization, which can be either accidental or intentional (Harrison and Larson, 2014). Hybridizing the two species, blue and Nile tilapia, in this way, is expected to have resulted in the desired higher salinity tolerance because blue tilapia has greater salinity tolerance than Nile tilapia (Nugon, 2003). Several studies have focused on identifying major QTLs

associated with salinity tolerance in tilapia and investigated how genomic architecture was shaped by either selection or introgression. However, the mechanism by which selection leads to tolerance to high salinity remains unknown. Using genome-scale sequence data provides the opportunity to detect signatures of introgression and selection. In this study, we aimed to understand the suspected origins of a domesticated tilapia strain, especially the genomic contributions from blue and Nile tilapia. Secondly, because the strain was selected for growth in high salinity over four generations, the role of introgression and selection in salinity tolerance in the Sukamandi strain was explored. Our study provides fundamental knowledge of a hybrid tilapia adapting to high salinity stress and provides the candidate genes in tilapia.

4.2 Methods

4.2.1 Ethics statement

The experiment was performed under a signed cooperation agreement between Wageningen University & Research, the Netherlands, and the Ministry of Marine Affairs and Fisheries, Indonesia on July 5, 2018. Fin clip samples of the Sukamandi strain were shipped following the Nagoya protocol which has been stipulated in the Material Transfer Agreement between RIFB and WUR on May 23, 2019 (permit ID: 1417/BRSDM/VIII/2019). All parties agreed to this experiment.

4.2.2 Sample collection and whole genome sequencing

We obtained fin clips from 20 fish of the Sukamandi strain from the RIFB (Subang, Java) in 2019. DNA for whole genome resequencing was isolated from fin clips using EchoLUTION Tissue DNA Micro Kit (BioECHO # 010 Na+/K+-ATPase-002-050) following company's specifications. DNA yield and quality were checked using a NanoDrop 2000 spectrophotometer (Thermo Scientific) and Qubit 2.0 Fluorometer (Invitrogen, Carlsbad, CA, USA). Qubit measurements of concentrations of the selected samples ranged from 40 to 100 ng/ul. The Nanodrop measurements of 260/280 ratios were between 1.7 and 1.9 while the 260/230 ratios were between 1.8 and 2.0. The sequencing libraries were prepared according to Illumina TruSeq HT protocol by a commercial provider (Novogene UK). Paired-end 150 bp sequencing strategy was performed for each library on an Illumina HiSeq 2500 machine (Illumina, USA) according to the manufacturers standard protocol.

The origin of the Sukamandi strain is in doubt, but potentially includes blue tilapia and Nile tilapia. Therefore, we downloaded publicly available blue tilapia genome data (accession number PRJNA358089 and PRJEB23203 from NCBI Sequence Read Archive). The blue tilapia individuals used were F1 individuals derived from a stock

originally provided by Dr. Gideon Hulata (Institute of Animal Science, Agricultural Research Organization, The Volcani Center, Bet Dagan, Israel) and maintained at University of Maryland (Conte et al., 2017). The Nile tilapia genome data were shared by aquaculture breeding company Genomar. The Nile tilapia is the Genomar supreme tilapia strain that originated from the GIFT population. The strain was selected for growth from generations 1 to 14, and for growth and filet yield from generations 15 to 19. The samples in this study were collected from generation 20. For more details see Joshi et al. (2018). To avoid bias between the newly generated Sukamandi strain data in the present study and the publicly available data, raw sequence data from representative individuals (four blue tilapia and seven Nile tilapia) with genomic coverage similar to the average of all Sukamandi samples were used for further analyses.

4.2.3 Read mapping and SNP identification

Quality control and trimming of raw sequence reads were conducted using trimmomatic (Bolger et al., 2014). Remaining adapter sequences and low-quality sequences (mean quality <20) were removed. Reads trimmed to fewer than 36 bases were removed entirely, and only read pairs for which both reads passed the quality criteria were retained for further processing.

Filtered reads from each sample were mapped to the tilapia (*Oreochromis niloticus*) reference genome (O_niloticus_UMD_NMBU) (Conte et al., 2017) using bwa -mem (Li and Durbin, 2009). The quality of the mapping results were evaluated with qualimap (García-Alcalde et al., 2012). Four low-quality Sukamandi strain samples, showing uneven coverage across the genome, were excluded from further analysis. Prior to SNP calling, samtools (Li et al., 2009) was used for sorting, merging, and removing potential PCR duplicates. To obtain high-quality SNPs for each of three distinct groups (blue tilapia, Nile tilapia and the Sukamandi strain), a joint-calling method was implemented with a haplotype-based variant detection in FreeBayes (Garrison and Marth, 2012) with default settings except for the parameters: --min-base-quality 20, --haplotype-length 0 and --ploidy 2. The preliminary SNP dataset was further filtered by vcftools (Danecek et al., 2011) with the following inclusion criteria: (1) genotype rate higher than 20; (2) a minimum read depth of 10X and a maximum read depth 80X; (3) heterozygous or non-reference within at least one individual; (4) indels removed. In total 19.6 million SNPs for 27 samples from three strains were identified and used for subsequent analyses.

4.2.4 Population structure analysis

Firstly, we generated a SNP dataset based on filtering for minor allele frequency (MAF) >0.05 and overall call rate >90% for 22 linkage groups. Secondly, in order to reduce the effect of ascertainment bias, we pruned SNPs based on Linkage Disequilibrium (LD) with a threshold of $r^2 > 0.2$ in windows of 50 SNPs by applying the command *--indep-pairwise* (50 10 0.2) in PLINK (Purcell et al., 2007). To investigate population structure, a principal component analysis (PCA) was performed using the *-pca* command in PLINK. A rootless neighbor-joining tree was constructed with Tree and Reticulogram Reconstruction (T-REX) (Boc et al., 2012) on the basis of the identity-by state similarity matrix calculated by PLINK. The resulting neighbor-joining tree was visualized in Figtree (http://tree.bio.ed.ac.uk/software/figtree/). The proportion of ancestry, and potential introgression, across three species was quantified using the Admixture software (Alexander et al., 2009).

4.2.5 Genetic differentiation (*Fst*)

To obtain a quantitative measure of genome-wide differentiation between the Sukamandi and the blue and Nile tilapia populations, fixation index (*Fst*) was calculated using vcftools, applying the -weir-fst-pop option (Danecek et al., 2011) with a sliding window approach (20-kb non-overlapping windows). The *Fst* per window was Z-transformed (z*Fst*) based on the formula:

$$zFst = \frac{(Fst \text{ per window} - \text{mean } Fst \text{ of all windows})}{\text{standard deviation of all windows } Fst} \quad (1)$$

The salinity level varied from 30 to 58 ppt during selective breeding in Indonesia, and we hypothesized that the key genes that make the Sukamandi strain survive and grow well in brackish water are derived from blue tilapia. Therefore, by pairwise comparison between the Sukamandi strain and blue tilapia, a zero value for *Fst*, indicating no genetic difference between the two populations, was considered to identify candidate salinity tolerance regions. The identical region between Nile tilapia and blue tilapia were removed since they are less relevant in this study. Assuming that the Sukamandi strain originated from hybridization between blue tilapia and Nile tilapia (as is confirmed in this paper), highly differentiated regions compared to Nile tilapia were considered potentially enriched for genes conferring salinity tolerance, derived from blue tilapia. Therefore, for the pairwise comparison between the Sukamandi strain and Nile tilapia, the genome-wide top 1% of genetically differentiated regions were used for further analysis.

4.2.6 Genetic diversity analyses

The frequencies of SNP variants in the Sukamandi strain were classified into four different categories based on minor allele frequency (MAF) according to (Peñaloza

et al., 2020): Common (MAF>0.3); Intermediate (0.1<MAF<0.3); Low (0.05<MAF <0.1); Rare (MAF<0.05). Expected heterozygosity (He), observed heterozygosity (Ho), and LD were calculated to evaluate the genetic diversity of the Sukamandi strain. Pair-wise SNPs within 300 kb were evaluated by the squared correlation coefficient (r^2) with default parameters with PopLDdecay program (Zhang et al., 2019).

Inbreeding usually results in increased homozygosity. One way of estimating the inbreeding coefficient is to calculate the proportion of the homozygous regions in all linkage groups, so called Runs of Homozygosity (ROH). The ROH-based inbreeding coefficient (F_{ROH}) was measured by the following:

$$\boldsymbol{F_{ROH}} = \frac{\sum \boldsymbol{L_{ROH}}}{\boldsymbol{L_{genome}}} \qquad (2)$$

where $\sum \boldsymbol{L_{ROH}}$ is the total length of homozygous regions of an individual and $\boldsymbol{L_{genome}}$ is the total genome length across the genome covered by SNPs (McQuillan et al., 2008). ROH analyses were performed using the -homozyg function in Plink. As pruning for LD can cause problems for ROH detection (Meyermans et al., 2020), the following parameters were applied in ROH segment identification with unpruned variant file of the Sukamandi strain, blue tilapia and Nile tilapia: (1) a sliding window of 50 SNPs across the genome; (2) the minimum ROH size was set to 100 kb; (3) allowed the windows to contain ≤2 missing genotype call but no heterozygous genotype; (4) the maximum interval size between two adjacent SNPs was 3000 kb. (5) proportion of homozygous overlapping windows was 0.05.

4.2.7 Artificial selection signatures and identity by descent (IBD) detection

To detect candidate regions showing a signature of artificial selection, we applied a site frequency spectrum (SFS) based method. Specifically, a composite likelihood ratio (CLR) test was performed to compare the allele frequency spectrum in candidate regions, as implemented in SweeD (Pavlidis et al., 2013). All CLR tests were run per linkage group with a resolution of 10 kb/window. Strong selection windows were defined as top 1% of the empirical distribution of the CLR scores. In addition, a different method to detect selection signals based on a distortion in allele frequency with Tajima' D calculation was also applied.

To examine the genetic contributions from blue tilapia and Nile tilapia to the Sukamandi strain, genomic regions showing potential introgression signatures were detected by calculating the frequency of shared Identity By Descent (IBD) (Bosse et

al., 2014). Specifically, we phased haplotypes for each linkage group separately in Beagle (Browning et al., 2018) with a sliding window size of 0.01 cM using the genetic map for the latest O_niloticus_UMD_NMBU assembly from Joshi et al. (2018). The number of iterations used to estimate genotype phase was 12. We set criteria for refining IBD as follows: windows = 0.03 cM, trim = 0.0001 cM, length = 0.01 cM and LOD = 3. To infer the relative fraction of haplotype sharing between the Sukamandi strain and blue tilapia, the counts of IBD (cIBD) segments were recorded per bin between the Sukamandi strain and blue tilapia, and then normalized for the total pair-wise comparison (tIBD). The normalized IBD (nIBD) between the Sukamandi strain and blue tilapia was then compared with nIBD computed between the Sukamandi strain and Nile tilapia. The relative values of nIBD were calculated according to Wu et al. (2020). Afterwards, if relative identity-by-descent (rIBD) > 0, this was regarded as a candidate introgression event from blue tilapia to the Sukamandi strain.

4.2.8 Functional annotation of candidate genes

Genomic regions that were identified as harbouring signatures of selection or introgression, were functionally annotated based on the latest tilapia reference genome (Conte et al., 2017). To better understand the function of tilapia genes, we first searched Ensembl genome browser (release 106) for the presence of orthologous genes. Then the amino acid sequences of those genes were compared using BLASTP (E-value threshold of 1e-6) against the protein sequences of the well-annotated zebrafish (*Danio rerio*) based on genome build GCF_000002035.6_GRCz11 (Yasuike et al., 2018). Subsequently, gene identifiers and functional annotation of the zebra fish genes were retrieved from the Ensembl Biomart database (Kinsella et al., 2011). Enrichment for Gene Ontology (GO) terms was then conducted using the Metascape database (Tripathi et al., 2015) by applying the zebrafish orthologs. Enriched terms with a minimum 3 gene overlap with the input gene list, p-value <0.05 and enrichment factor > 1.5 (the ratio between the observed counts and the counts expected by chance) were considered, which were visualized with the ggplot2 R package (https://github.com/tidyverse/ggplot2). Poisson distribution test was analyzed in R (version 4.0.2).

4.3 Results

4.3.1 Genetic diversity and population structure

In total 439.51 Gb raw Illumina sequence data were generated, resulting in an average of 75% of the genome being covered above 10X for each of the Sukamandi strain samples. For the reference Nile and blue tilapia samples, we included only

samples with over 70% of the genome and a minimum coverage of 10X. Because a high mapping rate (> 98%) to the Nile tilapia reference genome across all samples of the three species, we applied a joint-calling method and generated 19,620,053 SNPs, of which 19,446,400 were retained after quality control. Of these SNPs, only 16,902,167 were located on the 22 linkage groups (LGs) and were used for further analyses, 5,589,805 in blue tilapia, 7,402,876 in Nile tilapia and 12,545,465 in the Sukamandi strain. The genomic distribution of the SNPs was uniform except for LG3. The genomic region between 36.8 Mb and 80 Mb on LG3 showed an elevated SNP density (Supplementary Fig. 1).

To investigate the genetic relationships among individuals of the three species, we performed a PCA based on the genotypes of SNPs located on the 22 LGs. The first two components accounted for 18.2% and 13.7% of the genetic variation, respectively (Fig. 1a). The Sukamandi strain and Nile tilapia show the closest genetic relationship. Overall, principal components plots show unique clusters for all three strains, but, within groups, the Sukamandi samples showed the highest genetic variation.

The ADMIXTURE analysis indicates that the overall shared ancestry of Sukamandi strain with blue tilapia is 0.115 varying from 0.015 and 0.230 for K = 2 (Fig. 1b) and a detail of each individual present in Supplementary Table 3. Even at K = 3, blue tilapia ancestry is still inferred for the Sukamandi strain at proportions between 0 and 0.181. In the neighbor-joining tree (Fig. 1c), we see that Sukamandi samples clustered at a position intermediate between Nile tilapia and blue tilapia. Among the Sukamandi samples, four samples (Tilapia_1114_M, Tilapia_542_F, Tilapia_735_F, Tilapia_725_F) showed a closer genetic relationship with blue tilapia individuals, which also show high blue tilapia ancestry proportion in ADMIXTURE analyse. This suggests that hybridization was not uniform in the Sukamandi strain. Additionally, the lowest cross validation error was observed for admixture when K is equal to 3 (Supplementary Fig. 2).

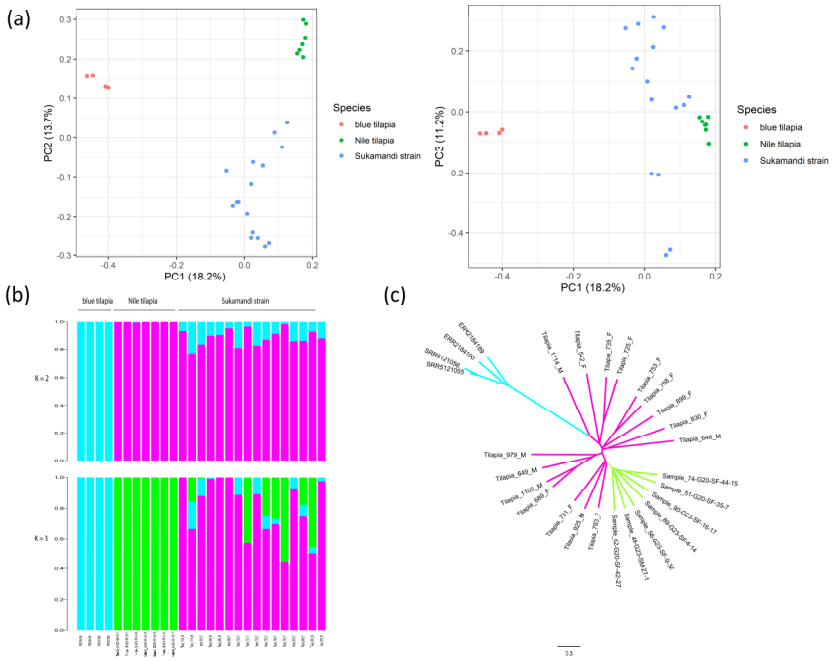


Fig 1. Genetic structure and phylogenetic relationship of blue and Nile tilapia tested in this study. (a) Principal component analysis of blue tilapia, Nile tilapia and Sukamandi strain. (b) Population structure of blue tilapia, Nile tilapia and Sukamandi strain analyzed by ADMIXTURE. The K value represents the number of assumed ancestries. (c) neighbor-joining tree based on an 1-ibs matrix between tilapia samples.

4.3.2 LD and ROH estimation for the Sukamandi strain

The minor allele frequency distribution of SNPs in the Sukamandi strain closely follows a Poison distribution (Probability = 0.987) (Fig 2a). The majority of MAF belongs to intermediate frequencies (0.1– 0.3), which includes 38.8% of all SNPs. The second MAF group is the low frequency bin (0.05–0.1) which includes 28.1% of the SNPs. These results indicated the majority SNPs are not fixed with the current generation. The LD was expressed as a function unit based on non-random association of alleles at different loci. The decay of LD at increasing physical distance between SNP pairs is presented in Fig 2b. A substantial drop in LD was observed for SNPs at distances from 0 to 20 kb.

The distribution of ROHs for the Sukamandi samples over 22 LGs is shown in Fig 2c. These individuals showed a variation in the content of ROH for different LGs, while LG19 showed the highest mean $\boldsymbol{F_{ROH}}$ (0.12) at LG level. The average length of ROHs in all samples was 157.41 kb, and average F_{ROH} with all individuals was 0.027

(Supplementary Table 1). A few samples, such as Tilapia_689_F, Tilapia_1105_M, Tilapia_899_F, showed a markedly higher content of ROHs. By comparing the runs of homozygosity to blue and Nile tilapia, the mean overall $\boldsymbol{F_{ROH}}$ is 0.027, which is between blue tilapia (0.037) and Nile tilapia (0.013). Overall, the Sukamandi strain was maintained at a low inbreeding level.

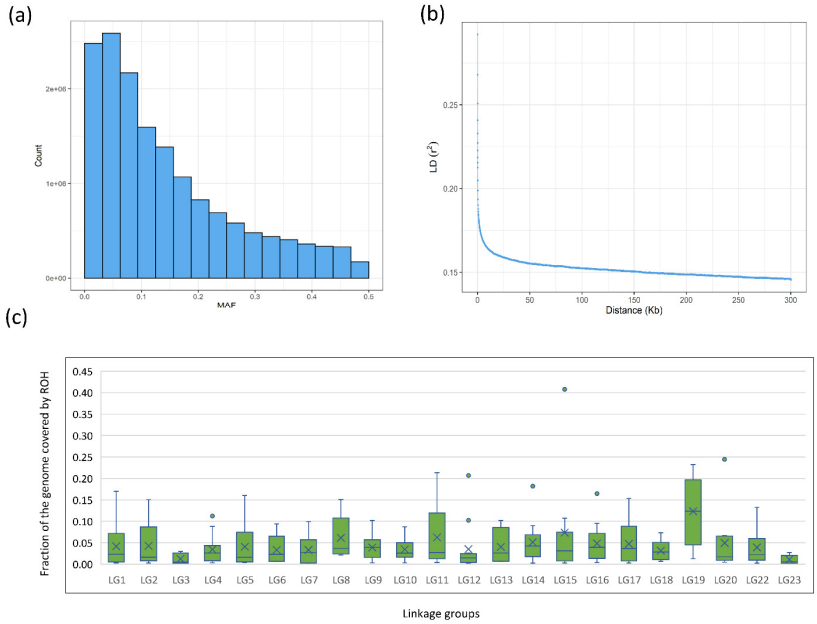


Fig 2. Minor allele frequency and inbreeding estimation for the Sukamandi strain. (a) minor allele frequency distribution of all biallelic SNPs in the Sukamandi strain. (b) LD decay with distance across the 22 linkage groups. (c) The total length of ROHs detected for each individual across the genome. Y axis denotes the fraction of total length of ROHs to genome length.

4.3.3 *Fst* and over-represented pathways related to genetic candidate regions

In order to understand why the Sukamandi strain performs well under high salinity conditions, two genetic mechanisms were inferred - selection and introgression. The rationale behind this approach is that the Sukamandi strain has been selected for growth in a high salinity environment for at least four generations since the beginning of the breeding programme. In addition, salinity tolerance was hypothesized to be derived, at least in part from one of the hybridized species - most likely blue tilapia. *Fst* is a measurement of population differentiation, and we applied *Fst* to identify regions that may have been selectively introgressed into the Sukamandi strain. The criteria for screening candidate regions with 20 kb per window included two parts: (1) genome-wide top 1% significantly differentiated regions in *zFst* results between Sukamandi and Nile tilapia strain (Fig 3a); (2) genetically identical regions (*zFst* = −1.97, equal

to *Fst* = 0) between Sukamandi and blue tilapia (Fig 3b). With these criteria, 449 candidate genomic regions containing 457 genes were identified by comparing Sukamandi and Nile tilapia. In the comparison between Sukamandi and blue tilapia, 68 candidate genomic regions containing 88 genes were identified. Only genomic interval (LG16:6740001 − 6760000, including four protein-coding genes *LOC102082836*, *LOC100690507*, *hibch*, *LOC102082489*), was found to be overlapping between those two sets of candidate genomic regions.

Gene Ontology (GO) enrichment analysis results for these two sets of genes are shown in Fig 4a and Fig 4b respectively. Over-represented terms in these 449 regions that differ from Nile tilapia include MAPK3 activity (GO:0004709, P < 0.01), potassium ion homeostasis (GO:0055075, P = 0.018), ATPase activity, coupled to transmembrane movement of ions (GO:0015662, P = 0.032), and response to calcium ion (GO:0051592, P = 0.034). Ion transport related terms were also enriched with 68 candidate regions between Sukamandi and blue tilapia, such as calcium ion binding (GO:0005509, P = 0.015) and regulation of ion transmembrane transport (GO:0034765, P = 0.025). Overall, 33 genes were identified in pathways relevant to osmoregulation (a detail of these genes are listed in Supplementary Table 2).

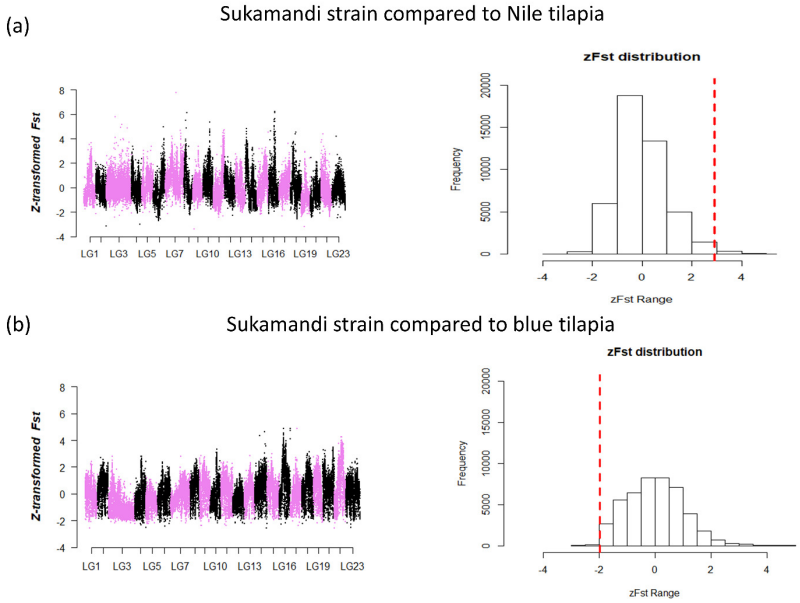


Fig 3. Pairwise genomic comparison for three tilapia species. (a) Mahattan Plot of z*Fst* value of Sukamandi strain versus Nile tilapia, the red dash line from histogram is top 1% cutoff at the whole genome level. (b) Mahattan Plot of *zFst* value of Sukamandi strain versus blue tilapia, the dashed red line from histogram is equivalent to *Fst* = 0.

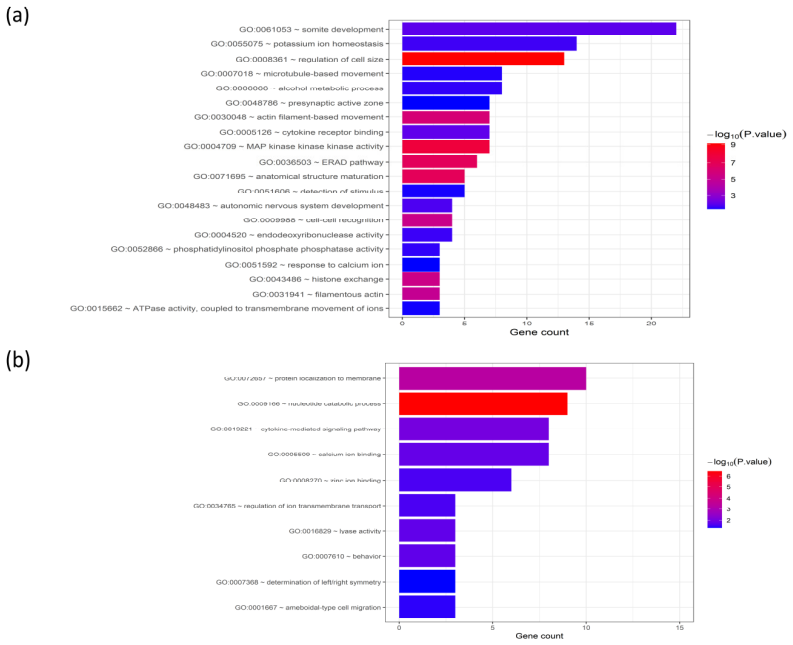


Fig 4. Functional enrichment for genetic candidate regions. (a) GO enrichment for differential regions in the comparison of the Sukamandi strain and Nile tilapia. (b) GO enrichment for identical regions in the comparison of the Sukamandi strain and blue tilapia.

4.3.4 Evidence for selection and introgression signatures from composite likelihood and IBD

Artificial selection was identified by scanning the genome using a composite likelihood ratio test (Fig 5). The highest number of candidate regions (10 kb/bin) under strong selection (top 1%) was identified on linkage group 1 (167, containing 428 protein-coding genes), and the lowest number (6) was found on linkage groups 3 (containing 54 protein-coding genes) and 11 (containing 58 protein-coding genes). Signal windows of top 1% selection overlapped genes containing, such as *caprin1a*, *nucb2a, abcb10* (LG1), *slc12a10.1* (LG3), *cacna1ab* (LG6), *ulk2* (LG14), *slc25a24* and *cdh1* (LG18). In addition, the genes *map3k8*, *mhc2dab* (LG3), *zgc:153039* (LG14), *slc9a2* (LG16) and *f2* (LG23) were also identified (top 5%). Although a strong selection peak was found on LG2, no gene was identified in that region with a known relation to osmoregulation.

The fraction of the blue tilapia genome shared with the Sukamandi strain is shown as rIBD, ranging from 0.27 (where 1 indicates all haplotypes are shared with blue tilapia and none with Nile tilapia origin) to −0.3 (a value of −1 indicates all haplotypes

are Nile tilapia) (Fig 6a). In total, 8.62% of the whole genome showed a positive value for rIBD (Fig 6b), which indicated potentially introgression from blue tilapia. The results indicate that the genome of the Sukamandi strain is overwhelming of Nile tilapia origin, consistent with the genetic structure, ADMIXTURE analysis, and neighbor-joining tree. Although we did not find strong overlap of introgression signatures with signatures of selection, several putative introgression haplotypes originating from blue tilapia (rIBD ranging from 0.001 to 0.06) were found to include potential salinity tolerance genes, including genes *slc25a24*, *slc12a10.1*, *zgc:153039*, *slc9a2*, *cdh1*.

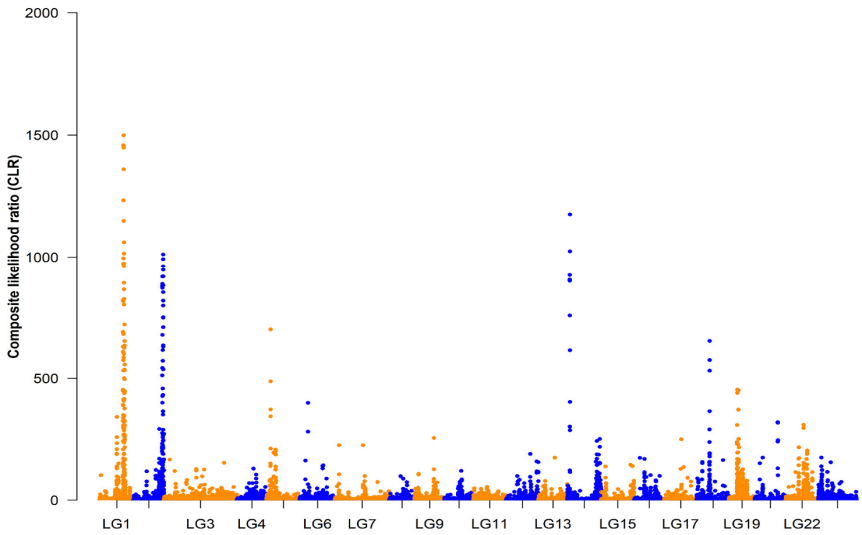


Fig 5. Selective sweep analysis for each linkage group. The x-axis denotes the linkage groups, and the y-axis shows the composite likelihood ratio evaluated with SweeD.

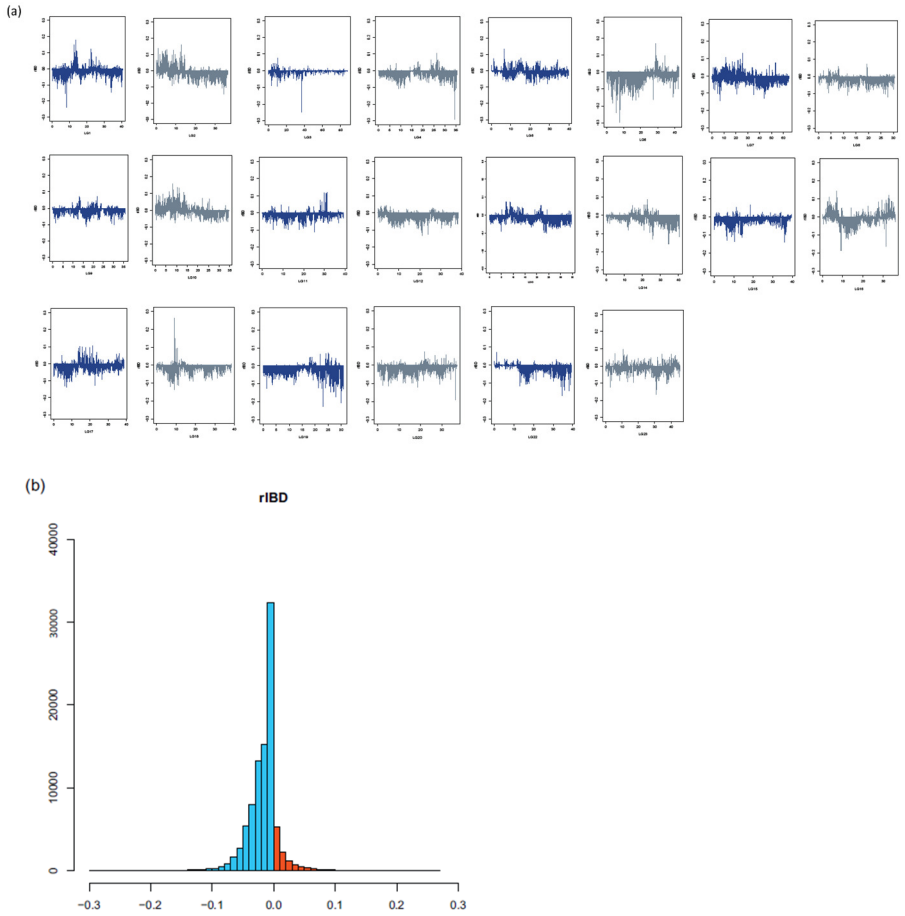


Fig 6. Genome-wide pattern of rIBD scores indicating introgressed haplotypes from blue tilapia to Sukamandi strain. (a) The rIBD score for 22 linkage groups: the positive values represent the haplotypes that are more genetically similar to blue tilapia, while the negative scores show the haplotypes that are more similar to Nile tilapia. (b) The distribution of rIBD scores for all linkages groups combined.

4.3.5 Potential molecular mechanism of adaption to high salinity

A central question addressed in this study is how species hybridization in combination with artificial selection may drive adaptation. By applying analyses to identify both signatures of selection and signatures of introgression, in total 457 genes derived from Nile tilapia and 88 genes from blue tilapia were identified. Based on GO and KEGG pathway annotations, 33 of these genes are involved in ion transport activity. Genes *slc25a24* and *cdh1* located between 13 and 14 Mb on LG18, are especially noteworthy because this region showed the highest values of rIBD

among the regions with genes involved in calcium ion activity. Besides, these two genes were identified based on two different pairwise comparisons of the Sukamandi strain with blue and Nile tilapia, respectively. *Slc25a24* was detected from the comparison between the Sukamandi strain and Nile tilapia within the top 1% of differences over the whole genome (Fig 7a), while *cdh1* was in a region where the Sukamandi strain was identical compared to blue tilapia (Fig 7b). Those two introgression regions showed a strong negative value of Tajima's D (Fig 7c), consistent with selection, and CLR indicated a weak selection signature for gene *slc25a24,* but strong selection for gene *cdh1* (Fig 7d). Interestingly, both genes showed evidence of introgression in the Sukamandi strain from blue tilapia.

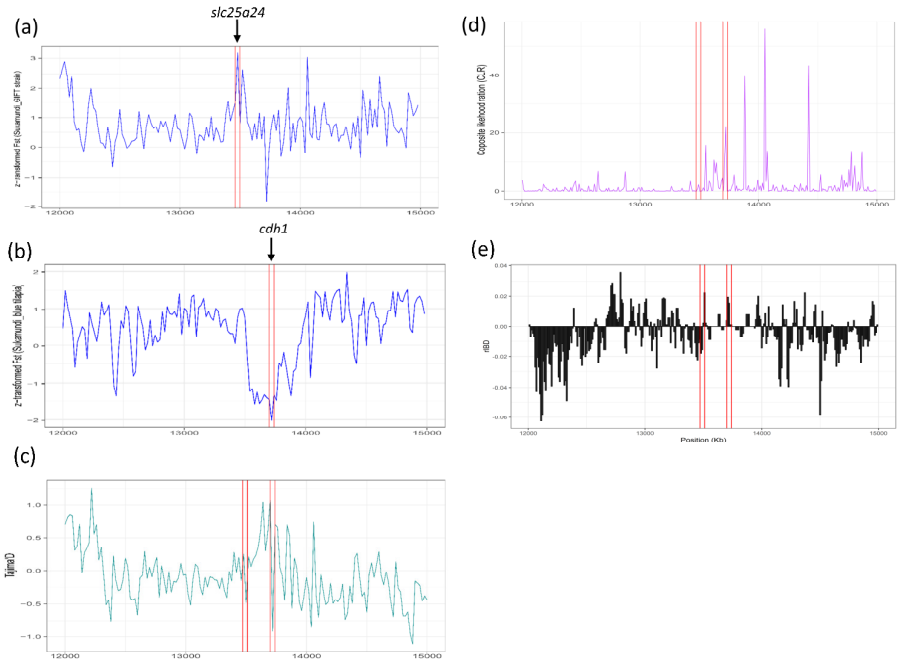


Fig 7. Plot of *zFst* (*(a) zFst* between the Sukamandi strain and Nile tilapia for the genomic region containing gene *slc25a24*; (b) *zFst* between the Sukamandi strain and blue tilapia for the genomic region gene *cdh1*), Tajima'D (c), composite likelihood ratio (CLR) (d) and rIBD (e) parameters for the region on LG18 including *slc25a24* and *cdh1*.

4.4 Discussion

4.4.1 The origin of the putative hybrid Sukamandi strain

RIBP initially believed that the Sukamandi strain obtained from a private aquaculture company in 2007 was blue tilapia (*Oreochromis aureus*). After mating with Nile tilapia at RIBP, the sex ratio clearly indicated that the Sukamandi strain is not pure

blue tilapia (unpublished result). We speculated that inadvertent mixing between blue tilapia and Nile tilapia occurred between 2008 and 2011 before the breeding program setup. Our results show clear distinction between the Sukamandi strain, Nile tilapia and blue tilapia for both the principal component and the admixture analysis (K = 3). We further compared the genome of the Sukamandi strain to that of blue tilapia and Nile tilapia to identify the functions of the candidate regions. Based on the genetic structure, the Sukamandi strain clearly represents a distinct population, while genetically closer to Nile tilapia than to blue tilapia. Furthermore, signatures of introgression from blue tilapia to the Sukamandi strain were clearly identified, which indicates that a blue tilapia x Nile tilapia cross was likely backcrossed with Nile tilapia for several generations. The Sukamandi strain, therefore, represents an interesting inter-species fish hybrid produced for aquaculture, but with extremely asymmetrical contributions from its parental populations.

In addition, we also observed that more SNPs on LG3 for the Sukamandi strain compared to blue and Nile tilapia, which might relate to species hybridization. LG3 is thought to be the sex chromosome in both blue tilapia (Lee et al., 2004) and Nile tilapia (Triay et al., 2020). Blue tilapia has a ZW/ZZ sex-determination system while Nile tilapia has a XX/XY sex-determination system. LG3 is the most repetitive chromosome in Nile tilapia. A wide region of sex-pattern differentiation was reported to be located on LG3 from ~40 to 85 Mb, with an accumulation of repetitive sequences (Conte et al., 2019). This could lead to a larger genomic difference between parent species, and hence a higher variation for these sex-determining loci in the hybrid population.

4.4.2 Over-represented pathways related to osmoregulation from genetic candidate regions

The hybridization between Nile tilapia and blue tilapia was able to create a more salinity tolerant, fast-growing strain. This suggests that we can find blue tilapia contributions to the Sukamandi strain to be over-represented with genes involved in osmoregulation. Conversely, we expect fewer contributions from Nile tilapia in osmoregulatory genes.

As expected, we identified four over-represented pathways potentially associated with salinity tolerance which included Mitogen-activated Protein kinase kinase kinase (MAPK3) activity, potassium ion homeostasis, ATPase activity (coupled to transmembrane movement of ions) and response to calcium ion. In addition, two other over-represented pathways involved in regulation of ion transmembrane

transport and calcium ion binding were discovered from genetic regions similar in both the Sukamandi strain and blue tilapia. Mitogen-activated protein kinases are members of a dynamic protein kinase network (Cuevas et al., 2007). Atay et al. (2017) showed that osmotic stress could induce MAPK3 activity. The action of ATPase, especially regarding Na+/K+ ion pumps, is critical for ion homeostasis for all teleost fish despite their habitat environment (Qin et al., 2020). The Na+/K+-ATPase regulation is regarded to be different during seawater and fresh water acclimation (Bystriansky et al., 2006). Two other enriched pathways (GO:0051592 response to calcium ion, GO:0005509 calcium ion binding) are also regarded as important pathways in adaptation to high salinity. It is known that Ca2+ absorption occurs in epithelia of many tissues, including kidney, intestine and gills (Hoenderop et al., 2005). Overall, 33 candidate salinity tolerance genes enriched in osmoregulation pathways were identified, giving insight into the genomic architecture of salinity tolerance in tilapia.

Comparing these 33 salinity tolerance genes to the previous tilapia QTL mapping studies related to salinity tolerance (Rengmark et al., 2007; Gu et al., 2018a, 2018b; Jiang et al., 2019), reveals almost no overlap with these salinity tolerance genes. But there are several previous findings showing that salinity tolerance genes similar to those identified in our study, differed in expression during a salinity challenge. For instance, *slc6a15* codes for solute carrier family 6 member 15, with a 1-to-1 ortholog in many fish species such as European seabass (*Dicentrarchus labrax*), zebrafish (*Danio rerio*), and Japanese medaka (*Oryzias latipes*). This gene was strikingly up-regulated (28 fold change) and reported as a key gene in liver of spotted seabass (*Dicentrarchus punctatus*) during salinity adaptation (Zhang et al., 2017). Ivanis et al. (2008) suggest that gene *slc9a2* regulates Na+/H+ exchange in freshwater rainbow trout (*Oncorhynchus mykiss*). This gene was also reported as a 1-to-1 ortholog between Nile tilapia and rainbow trout on Ensembl genome browser (release 106). Gene *slc25a24* has been shown to be involved in calcium regulation and to be differentially expressed under stress by comparative transcriptomic analysis in turbot (*Scophthalmus maximus*) (Cui et al., 2020). Gene *slc4a7* showed induced expression levels after transferring from seawater to freshwater in spotted sea bass (*Lateolabrax maculatus*) (Wang et al., 2020). And Na+/HCO3 cotransporter encoded by *SLC4A7* is widely distributed in epithelial and nonepithelial tissues (Aalkjaer et al., 2014). Chen et al. (2018) observed that the high expression of cadherin 1 (*cdh1*) in freshwater could reduce the paracellular ions loss from the plasma to the low ionic environment, while a relatively low expression in seawater facilitate the paracellular exist of Na^+ in the gills of climbing perch (*Anabas testudineus*), therefore maintaining ionic balance.

4.4.3 Genetic candidate regions driven by selection and introgression

In the Sukamandi strain, the over-represented pathways related to osmoregulation could be the result of selection, introgression or even genetic drift. In order to understand the main genetic driver for these regions, the impact of selection and introgression for salinity tolerance on the genome was further explored. Our *Fst* results suggest that genetically differentiated regions identified are partly due to selection, while rIBD values suggest some of these regions to be a consequence of introgression from blue tilapia, such as the region that includes the *slc25a24* and *cdh1* genes. It is often observed that there is a degree of inconsistency when comparing different measures of selection and introgression. This may be due to different aspects of selection being captured by any one of these analyses (Brawand et al., 2015). Furthermore, it is often observed that regions under selection lie in non-coding regions of the tilapia genome (Xia et al., 2015).

The Sukamandi strain has been selected for growth in brackish water ponds for four generations and this might account for the selection signatures observed. The introgression signals related to osmoregulation are maintained after four-generation selection. Martin et al. (2017) showed that selection on individual loci may become weak when combined with introgression, especially if the direction of selection is opposite to introgression. Genetically differentiated regions that result from introgression only, such as the genes *kcnh4*, *stac*, *slco3a1*, *cpne4a*, *atp2c1*, could have a negative effect in the high salinity environment, resulting in specific incompatibilities. On the other hand, it is very likely introgressed loci are retained by chance. Only four generations of selection have passed, which will not be sufficient to establish a uniform contribution of the parent populations in each of the current Sukamandi strain.

However, given that the breeding goals for the Sukamandi strain, explicitly salinity tolerance, it is most likely that calcium ion binding pathways were under selection over the past four generations of selections. Calcium regulation and homeostasis play an essential role in osmotic regulation in several fish species, such as the turbot (*Scophthalmus maximus*) (Cui et al., 2020) and green spotted puffer fish (*Tetraodon nigroviridis*) (Pinto et al., 2010). Both introgression and signature of selection analyses similarly suggest an important role for calcium regulation in the breeding for salinity tolerance in the Sukamandi strain.

4.5 Conclusions

The Sukamandi strain is genetically overwhelmingly of Nile tilapia origin, while introgression from blue tilapia has likely contributed to the observed salinity tolerance. Thirty-three salinity tolerance genes enriched in ion transmembrane transport processes, such as MAPK3 activity, potassium ion homeostasis, ATPase activity and response to calcium ion, explained the underlying mechanism of adaptation to high salinity environment. Selective breeding for growth and survival in high salinity indicates that specific genes from blue tilapia have been preferentially kept during that selection. These findings not only contribute to understanding salinity tolerance in tilapia, but also provide an interesting model for hybrid introgression in farmed species in general.

4.6 Funding

This study was financially supported for genome sequencing by the Koepon foundation (Project number 4164016900). We thank China Scholarship Council (CSC, File No.201803250069), for sponsoring Xiaofei Yu's PhD study at Wageningen University and Research. The funding bodies had no role in study design, data collection, analysis, and result interpretation in this study.

4.7 Data availability statement

Sequence information of the utilized samples is deposited in the European nucleotide archive (https://www.ebi.ac.uk/ena/browser/home) under accession number PRJEB50266.

4.8 Author contributions

X.Y: Writing – original draft, Methodology, Formal analysis, Investigation, Visualization. P.S: Writing – review & editing, Investigation, Data curation. J.W.M.B: Writing – review & editing, Investigation, Project administration. L.L: Writing – review & editing, Methodology, Investigation. I.I: Writing – review & editing, Conceptualization, Resources. M.A.M.G: Writing – review & editing, Supervision, Conceptualization, Methodology, Investigation. H.K: Writing – review & editing, Supervision, Conceptualization, Funding acquisition, Investigation. H.J.M: Writing – review & editing, Supervision, Conceptualization, Methodology, Investigation.

4.9 Ethics approval and consent to participate

All experiments of the current study were performed under a signed cooperation agreement between Wageningen University & Research (WUR) and Ministry of Marine Affairs and Fisheries on July 5, 2018. Fin clip samples of the Sukamandi strain were obtained following the Nagoya protocol which has been stipulated in the

Mutual Transfer Agreement (MTA) between RIFB and WUR on May 23, 2019 (permit ID: 1417/BRSDM/VIII/2019). All parties agreed to this experiment.

4.10 Declaration of Competing Interest

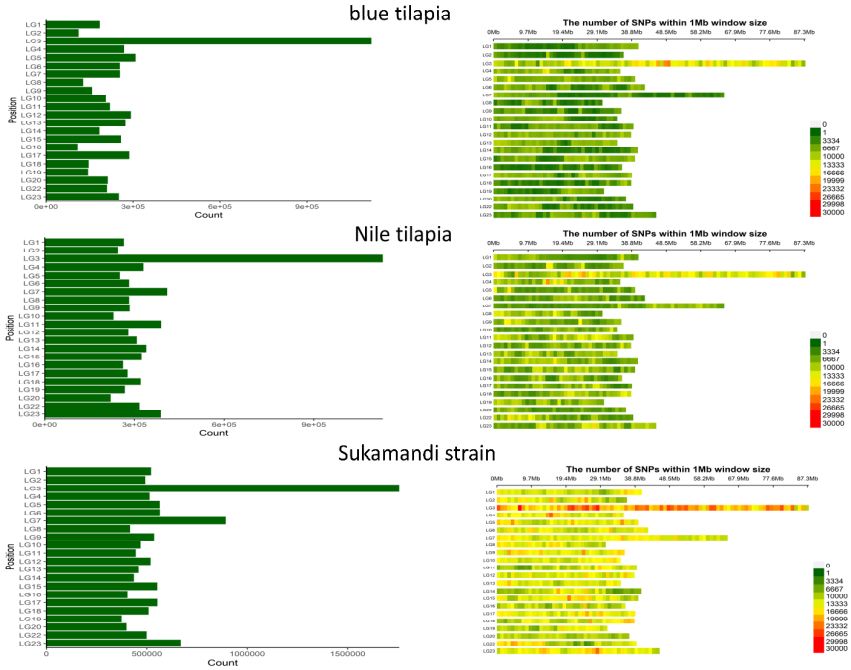
The authors declare that they have no commercial conflicts or competing interests that could influence the work reported here.

4.11 Acknowledgements

The authors thank colleagues from RIFB in Indonesia for assisting with the fin clips collection and shipment, and to Bert Dibbits and Kimberley Laport for DNA extraction and preparation of the sequencing libraries. Rajesh Joshi from GenoMar kindly provided a high-quality whole genome sequencing dataset of Nile tilapia. We thank Mirte Bosse, Rayner González-Prendes and Zhou Wu for the help in data analysis and Mark Camara for help in editing.

4.12 Supplementary Information

For a compact layout, I have not included all Supplementary material in the thesis, only those which may assist the reader. For more information, the complete Supplementary figures and tables of the article are available at https://doi.org/10.1016/j.aquaculture.2022.738527.



Supplementary Fig 1. Histogram of SNPs distribution across all linkage groups and SNP density plots across all linkage groups for three tilapia species.

Supplementary Fig 2. Cross validation error from admixture analysis.

Supplementary Fig 3. Runs of homozygosity analyses for blue tilapia, Nile tilapia and the Sukamandi strain.

Supplementary Table 1. Inbreeding coefficient (F_{ROH}) for all individual of the Sukamandi strain.

Supplementary Table 2. Overview of candidate genes involved in osmoregulation over the whole genome.

Supplementary Table 3. The shared components among three tilapia species.

References

Aalkjaer, C., Boedtkjer, E., Choi, I., Lee, S., 2014. Cation-coupled bicarbonate transporters. Comprehensive Physiology 4, 1605–1637. https://doi.org/10.1002/cphy.c130005

Alexander, D.H., Novembre, J., Lange, K., 2009. Fast model-based estimation of ancestry in unrelated individuals. Genome research 19, 1655–1664.

Ali, A., Al-Tobasei, R., Lourenco, D., Leeds, T., Kenney, B., Salem, M., 2020. Genome-wide scan for common variants associated with intramuscular fat and moisture content in rainbow trout. BMC genomics 21, 1–17.

Atay, O., Skotheim, J.M., 2017. Spatial and temporal signal processing and decision making by MAPK pathwaysDynamic signal processing by MAPK pathways. The Journal of cell biology 216, 317–330.

Boc, A., Diallo, A.B., Makarenkov, V., 2012. T-REX: a web server for inferring, validating and visualizing phylogenetic trees and networks. Nucleic Acids Research 40, W573–W579. https://doi.org/10.1093/nar/gks485.

Bolger, A.M., Lohse, M., Usadel, B., 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30, 2114–2120.

Bosse, M., Megens, H.J., Frantz, L.A.F., Madsen, O., Larson, G., Paudel, Y., Duijvesteijn, N., Harlizius, B., Hagemeijer, Y., Crooijmans, R.P.M.A., Groenen, M.A.M., 2014. Genomic analysis reveals selection for Asian genes in European pigs following human-mediated introgression. Nature Communications 5, 1–8. https://doi.org/10.1038/ncomms5392.

Brawand, D., Wagner, C.E., Li, Y.I., Malinsky, M., Keller, I., Fan, S., Di Palma, F., 2015. The genomic substrate for adaptive radiation in African cichlid fish. Nature 513, 375–381. https://doi.org/10.1038/nature13726. Breves, J.P., Serizier, S.B., Goffin, V., McCormick, S.D., Breves, J.P., Serizier, S.B., Goffin, V., McCormick, S.D., Karlstrom, R.O., 2013. Prolactin regulates transcription of the ion uptake Na+/Cl− cotransporter (ncc) gene in zebrafish gill. Molecular and cellular endocrinology 369, 98–106.

Browning, B.L., Zhou, Y., Browning, S.R., 2018. A one-penny imputed genome from next-generation reference panels. The American Journal of Human Genetics 103, 338–348.

Bystriansky, J.S., Richards, J.G., Schulte, P.M., Ballantyne, J.S., 2006. Reciprocal expression of gill Na+/K+-ATPaseα-subunit isoforms α1a and α1b during seawater acclimation of three salmonid fishes that vary in their salinity tolerance. Journal of Experimental Biology 209, 1848–1858.

Chen, X.L., Lui, E.Y., Ip, Y.K., Lam, S.H., 2018. RNA sequencing, de novo assembly and differential analysis of the gill transcriptome of freshwater climbing perch Anabas testudineus after 6 days of seawater exposure. Journal of Fish Biology 93, 215–228. https://doi.org/10.1111/jfb.13653.

Cnaani, A., Velan, A., Hulata, G., Liping, L., Fitzsimmons, K., 2011. Improving salinity tolerance in tilapias: A REVIEW, in: Better Science, Better Fish, Better Life. Proceedings of the Ninth International Symposium on Tilapia in Aquaculture.

Conte, M.A., Gammerdinger, W.J., Bartie, K.L., Penman, D.J., Kocher, T.D., 2017. A high quality assembly of the Nile Tilapia (Oreochromis niloticus) genome reveals the structure of two sex determination regions. BMC Genomics 18, 1–19. https://doi.org/10.1186/s12864-017-3723-5.

Conte, Matthew A., Joshi, Rajesh, Moore, Emily C, Nandamuri, Sri Pratima, Gammerdinger, William J., Roberts, Reade B., Carleton, Karen L., Lien, Sigbjørn, Kocher, Thomas D., 2019. Chromosome-scale assemblies reveal the structural evolution of African

cichlid genomes. GigaScience 8 (4), giz030. https://doi.org/10.1093/gigascience/giz030.

Cuevas, B.D., Abell, A.N., Johnson, G.L., 2007. Role of mitogen-activated protein kinase kinase kinases in signal integration. Oncogene 26, 3159–3171. https://doi.org/10.1038/sj.onc.1210409.

Cui, W., Ma, A., Huang, Z., Wang, X., Liu, Z., Xia, D., Yang, S., Zhao, T., 2020. Comparative transcriptomic analysis reveals mechanisms of divergence in osmotic regulation of the turbot Scophthalmus maximus. Fish Physiology and Biochemistry 46, 1519–1536. https://doi.org/10.1007/s10695-020-00808-6.

Danecek, P., Auton, A., Abecasis, G., Albers, C.A., Banks, E., DePristo, M.A., Handsaker, R.E., Lunter, G., Marth, G.T., Sherry, S.T., 2011. The variant call format and VCFtools. Bioinformatics 27, 2156–2158.

Deane, E.E., Woo, N.Y.S., 2009. Modulation of fish growth hormone levels by salinity, temperature, pollutants and aquaculture related stress: A review. Reviews in Fish Biology and Fisheries 19, 97–120. https://doi.org/10.1007/s11160-008-9091-0

El-Sayed, M., A. F., 2006. Tilapia culture in salt water: environmental requirements, nutritional implications and economic potentials. Avances en Nutrici´on Acuícola VIII 95–106.

El-Sayed, A.-F.M., 2019. Tilapia culture. Academic Press.

Fitzsimmons, K.M., Shahkar, E., 2017. Tilapia-shrimp polyculture. Tilapia in Intensive Co-culture, First Edition. Edited by Peter W. Perschbacher and Robert R. Stickney. John Wiley & Sons, Ltd 94–113.

García-Alcalde, F., Okonechnikov, K., Carbonell, J., Cruz, L.M., Götz, S., Tarazona, S., Dopazo, J., Meyer, T.F., Conesa, A., 2012. Qualimap: Evaluating next-generation sequencing alignment data. Bioinformatics 28, 2678–2679. https://doi.org/10.1093/bioinformatics/bts503.

Garrison, E., Marth, G., 2012. Haplotype-based variant detection from short-read sequencing. arXiv preprint arXiv:1207.3907.

Gu, X.H., Jiang, D.L., Huang, Y., Li, B.J., Chen, C.H., Lin, H.R., Xia, J.H., 2018a. Identifying a Major QTL Associated with Salinity Tolerance in Nile Tilapia Using QTL-Seq. Marine Biotechnology 20, 98–107. https://doi.org/10.1007/s10126-017-9790-4.

Gu, X.H., Lin, H.R., Xia, J.H., 2018b. Significant association of SNP polymorphism in the tilapia enhancer of polycomb homolog 1 gene with salt tolerance. Aquaculture Research 49, 1690–1698. https://doi.org/10.1111/are.13625.

Harrison, R.G., Larson, E.L., 2014. Hybridization, introgression, and the nature of species boundaries. Journal of Heredity 105, 795–809.

Hoenderop, J.G.J., Nilius, B., Bindels, R.J.M., 2005. Calcium absorption across epithelia. Physiological reviews 85, 373–422.

Ivanis, G., Esbaugh, A.J., Perry, S.F., 2008. Branchial expression and localization of SLC9A2 and SLC9A3 sodium/hydrogen exchangers and their possible role in acid-base regulation in freshwater rainbow trout (Oncorhynchus mykiss). Journal of Experimental Biology 211, 2467–2477. https://doi.org/10.1242/jeb.017491.

Jaspe, C.J., Caipang, C.M.A., 2011. Increasing salinity tolerance in tilapias: selective breeding using locally available strains. Aquaculture, Aquarium, Conservation & Legislation 4, 437–441.

Jiang, D.L., Gu, X.H., Li, B.J., Zhu, Z.X., Qin, H., Meng, Z. ning, Lin, H.R., Xia, J.H., 2019. Identifying a Long QTL Cluster Across chrLG18 Associated with Salt Tolerance in Tilapia Using GWAS and QTL-seq. Marine Biotechnology 21, 250–261. https://doi.org/10.1007/s10126-019-09877-y.

Joshi, R., Árnyasi, M., Lien, S., Gjøen, H.M., Alvarez, A.T., Kent, M.P., Tola Alvarez, A., Kent, M.P., 2018. Development and validation of 58K SNP-array and high-density linkage

map in Nile tilapia (O. niloticus). Frontiers in Genetics 9. https://doi.org/10.3389/fgene.2018.00472.

Junior, A.P.B., Azevedo, C.M. da S.B., Pontes, F.S.T., Henry-Silva, G.G., 2012. Polyculture of Nile tilapia and shrimp at different stocking densities. Revista Brasileira de Zootecnia 41, 1561–1569. https://doi.org/10.1590/S1516-35982012000700002.

Kalujnaia, S., McWilliam, I.S., Zaguinaiko, V.A., Feilen, A.L., Nicholson, J., Hazon, N., Cramb, G., 2007. Salinity adaptation and gene profiling analysis in the European eel (Anguilla anguilla) using microarray technology. Gen. Comp. Endocrinol. 152, 274–280. https://doi.org/10.1016/j.ygcen.2006.12.025.

Kinsella, R.J., Kähäri, A., Haider, S., Zamora, J., Proctor, G., Spudich, G., Almeida-King, J., Staines, D., Derwent, P., Kerhornou, A., 2011. Ensembl BioMarts: a hub for data retrieval across taxonomic space. Database 2011.

Kusakabe, M., Ishikawa, A., Ravinet, M., Yoshida, K., Makino, T., Toyoda, A., Fujiyama, A., Kitano, J., 2017. Genetic basis for variation in salinity tolerance between stickleback ecotypes. Molecular ecology 26, 304–319.

Laverty, G., Skadhauge, E., 2012. Adaptation of teleosts to very high salinity. Comparative Biochemistry and Physiology - A Molecular and Integrative Physiology 163, 1–6. https://doi.org/10.1016/j.cbpa.2012.05.203.

Lawson, E.O., Anetekhai, M.A., 2011. Salinity Tolerance and Preference of Hatchery Reared Nile Tilapia, Oreochromis niloticus (Linneaus 1758). Asian Journal of Agricultural Science 3, 104–110.

Le Bras, Y., Dechamp, N., Krieg, F., Filangi, O., Guyomard, R., Boussaha, M., Bovenhuis, H., Pottinger, T.G., Prunet, P., Le Roy, P., 2011. Detection of QTL with effects on osmoregulation capacities in the rainbow trout (Oncorhynchus mykiss). BMC genetics 12, 1–14.

Lee, B.Y., Hulata, G., Kocher, T.D., 2004. Two unlinked loci controlling the sex of blue tilapia (Oreochromis aureus). Heredity 92, 543–549.

Li, H., Durbin, R., 2009. Fast and accurate short read alignment with Burrows–Wheeler transform. bioinformatics 25, 1754–1760.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., 2009. The sequence alignment/map format and SAMtools. Bioinformatics 25, 2078–2079.

Martin, S.H., Jiggins, C.D., 2017. Interpreting the genomic landscape of introgression. Current opinion in genetics & development 47, 69–74.

Martínez-Porchas, M., Martínez-Córdova, L.R., Porchas-Cornejo, M.A., López-Elías, J.A., 2010. Shrimp polyculture: a potentially profitable, sustainable, but uncommon aquacultural practice. Reviews in Aquaculture 2, 73–85.

McCormick, S.D., Regish, A.M., Christensen, A.K., 2009. Distinct freshwater and seawater isoforms of Na+/K +-ATPase in gill chloride cells of Atlantic salmon. Journal of Experimental Biology 212, 3994–4001. https://doi.org/10.1242/jeb.037275.

McQuillan, R., Leutenegger, A.-L., Abdel-Rahman, R., Franklin, C.S., Pericic, M., Barac-Lauc, L., Smolej-Narancic, N., Janicijevic, B., Polasek, O., Tenesa, A., 2008. Runs of homozygosity in European populations. The American Journal of Human Genetics 83, 359–372.

Meyermans, R., Gorssen, W., Buys, N., Janssens, S., 2020. How to study runs of homozygosity using plink? a guide for analyzing medium density snp data in livestock and pet species. BMC Genomics 21, 1–14. https://doi.org/10.1186/s12864-020-6463-x.

Mjoun, K, Kurt, A., Brown, L.M., 2010b. Tilapia : Environmental biology and nutritional requirements. South Dakota Coopertaive Extension Service FS963-027.

Mjoun, Kamal, Rosentrater, K., Brown, M.L., 2010a. TILAPIA : Profile and Economic Importance Tilapia : Profile and Economic Importance.

Norman, J.D., Robinson, M., Glebe, B., Ferguson, M.M., Danzmann, R.G., 2012. Genomic arrangement of salinity tolerance QTLs in salmonids: A comparative analysis of Atlantic salmon (Salmo salar) with Arctic charr (Salvelinus alpinus) and rainbow trout (Oncorhynchus mykiss). BMC Genomics 13, 1–15. https://doi.org/10.1186/1471-2164-13-420.

Nugon, R.W., 2003. Salinity tolerance of juveniles of four varieties of tilapia.

Pavlidis, P., Živković, D., Stamatakis, A., Alachiotis, N., 2013. SweeD: Likelihood-based detection of selective sweeps in thousands of genomes. Molecular Biology and Evolution 30, 2224–2234. https://doi.org/10.1093/molbev/mst112.

Peñaloza, C., Robledo, D., Barría, A., Trinh, T.Q., Mahmuddin, M., Wiener, P., Benzie, J.A.H., Houston, R.D., 2020. Development and validation of an open access SNP array for Nile Tilapia (oreochromis niloticus). G3: Genes, Genomes, Genetics 10, 2777–2785. https://doi.org/10.1534/g3.120.401343.

Pinto, P.I.S., Matsumura, H., Thorne, M.A.S., Power, D.M., Terauchi, R., Reinhardt, R., Canário, A.V.M., 2010. Gill transcriptome response to changes in environmental calcium in the green spotted puffer fish. BMC Genomics 11. https://doi.org/10.1186/1471-2164-11-476.

Ponzoni, R.W., Nguyen, N.H., Khaw, H.L., Hamzah, A., Bakar, K.R.A., Yee, H.Y., 2011. Genetic improvement of Nile tilapia (Oreochromis niloticus) with special reference to the work conducted by the World Fish Center with the GIFT strain. Reviews in Aquaculture 3, 27–41. https://doi.org/10.1111/j.1753-5131.2010.01041.x.

Prabu, E., Rajagopalsamy, C.B.T., Ahilan, B., Jeevagan, I.J.M.A., Renuhadevi, M., 2019. Tilapia – An Excellent Candidate Species for World Aquaculture: A Review. Annual Research & Review in Biology 1–14. https://doi.org/10.9734/arrb/2019/v31i330052.

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A.R., Bender, D., Maller, J., Sklar, P., De Bakker, P.I.W., Daly, M.J., 2007. PLINK: a tool set for whole-genome association and population-based linkage analyses. The American journal of human genetics 81, 559–575.

Qin, H., Zhu, Z.X., Lin, H.R., Xia, J.H., Jia, Y.X., 2020. Exploring candidate genes in a major QTL region associated with salinity tolerance in the skin of Nile tilapia based on transcriptomic analysis. Aquaculture 526, 735380. https://doi.org/10.1016/j.aquaculture.2020.735380.

Rengmark, A. H., Slettan, A., Lee, W.J., Lie, Lingaas, F., 2007. Identification and mapping of genes associated with salt tolerance in tilapia. Journal of Fish Biology 71, 409–422. https://doi.org/10.1111/j.1095-8649.2007.01664.x.

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.

Sciara, A.A., Rodríguez-Ramilo, S.T., Hermida, M., Gómez-Tato, A., Fernández, J., Bouza, C., Martínez, P., 2018. Validation of growth-related quantitative trait loci markers in turbot (Scophthalmus maximus) families as a step toward marker assisted selection. Aquaculture 495, 602–610.

Setyawan, P., Robisalmi, A., 2014. PENGARUH SALINITAS TINGGI TERHADAP PERTUMBUHAN, SINTASAN, DAN HETEROSIS PADA PERSILANGAN IKAN NILA HITAM (Oreochromis niloticus) DAN IKAN NILA BIRU (Oreochromis aureus), in: Prosiding FORUM INOVASI TEKNOLOGI AKUAKULTUR. pp. 901–910.

Shapiro, B.J., Alm, E.J., 2008. Comparing patterns of natural selection across species using selective signatures. PLoS genetics 4, e23.

Triay, C., Conte, M.A., Baroiller, J.F., Bezault, E., Clark, F.E., Penman, D.J., Kocher, T.D., D'cotta, H., 2020. Structure and sequence of the sex determining locus in two wild populations of nile tilapia. Genes 11, 1–30. https://doi.org/10.3390/genes11091017

Tripathi, S., Pohl, M.O., Zhou, Y., Rodriguez-Frandsen, A., Wang, G., Stein, D.A., Moulton, H.M., DeJesus, P., Che, J., Mulder, L.C.F., 2015. Meta-and orthogonal integration of influenza "OMICs" data defines a role for UBR4 in virus budding. Cell host & microbe 18, 723–735.

Wang, L.Y., Tian, Y., Wen, H.S., Yu, P., Liu, Y., Qi, X., Gao, Z.C., Zhang, K.Q., Li, Y., 2020. SLC4 gene family in spotted sea bass (Lateolabrax maculatus): Structure, evolution, and expression profiling in response to alkalinity stress and salinity changes. Genes 11, 1–19. https://doi.org/10.3390/genes11111271.

Wu, Z., Bortoluzzi, C., Derks, M.F.L., Liu, L., Bosse, M., Hiemstra, S.J., Groenen, M.A.M., Crooijmans, R.P.M.A., 2020. Heterogeneity of a dwarf phenotype in Dutch traditional chicken breeds revealed by genomic analyses. Evolutionary Applications. https://doi.org/10.1111/eva.13183.

Xia, J.H., Bai, Z., Meng, Z., Zhang, Y., Wang, L., Liu, F., Jing, W., Wan, Z.Y., Li, J., Lin, H., 2015. Signatures of selection in tilapia revealed by whole genome resequencing. Scientific reports 5, 1–10.

Yan, J.J., Lee, Y.C., Tsou, Y.L., Tseng, Y.C., Hwang, P.P., 2020. Insulin-like growth factor 1 triggers salt secretion machinery in fish under acute salinity stress. Journal of Endocrinology 246, 277–288. https://doi.org/10.1530/JOE-20-0053.

Yasuike, M., Iwasaki, Y., Nishiki, I., Nakamura, Y., Matsuura, A., Yoshida, K., Noda, T., Andoh, T., Fujiwara, A., 2018. The yellowtail (Seriola quinqueradiata) genome and transcriptome atlas of the digestive tract. DNA Research 25, 547–560. https://doi.org/10.1093/dnares/dsy024.

Zak, T., Deshev, R., Benet-Perlberg, A., Naor, A., Magen, I., Shapira, Y., Ponzoni, R.W., Hulata, G., 2014. Genetic improvement of Israeli blue (Jordan) tilapia, Oreochromis aureus (Steindachner), through selective breeding for harvest weight. Aquaculture Research 45, 546–557. https://doi.org/10.1111/are.12072.

Zhang, C., Dong, S.-S., Xu, J.-Y., He, W.-M., Yang, T.-L., 2019. PopLDdecay: a fast and effective tool for linkage disequilibrium decay analysis based on variant call format files. Bioinformatics 35, 1786–1788.

Zhang, X., Wen, H., Wang, H., Ren, Y., Zhao, J., Li, Y., 2017. RNA-Seq analysis of salinity stress-responsive transcriptome in the liver of spotted sea bass (Lateolabrax maculatus). PLoS ONE 12. https://doi.org/10.1371/journal.pone.0173238.

5

Genome-wide association analyses reveal genotype-by-environment interactions for growth traits and organ weights in gilthead seabream (*Sparus aurata*)

Xiaofei Yu[1], John W.M. Bastiaansen[1], Benan Gulzari[1], Mark Camara[1], Han A. Mulder[1], Hans Komen[1], Martien A.M. Groenen[1], Hendrik-Jan Megens[1]

[1]Animal Breeding and Genomics, Wageningen University & Research, the Netherlands.

To be submitted.

Abstract

Background: Gilthead seabream is one of the most important fish species for Mediterranean aquaculture. More than 80% of the genetically improved fingerlings of this species originate from a single country, Greece. So far, QTL analysis has received some attention for growth, but other features, such as organ weights, have so far received little attention. A previous study has revealed moderate genotype-by-environment interactions for growth traits and organ weights between two distinct commercial production sites in Greece and Spain. The main objectives of this study were to identify QTLs and genes associated with growth traits and organ weights in gilthead seabream, and to identify the biological processes and pathways involved in these traits under different temperature environments.

Results: For growth traits, we found fourteen genome-wide suggestive SNP effects in Spain. A strong peak between 19.66 Mb to 46.84 Mb on chromosome 22 was shared among three growth traits (harvest weight, fillet weight and thermal growth coefficient). Fourteen genome-wide suggestive SNP effects were also identified in Greece. However, none of those 14 candidate SNPs overlap with results found in Spain. Two SNPs on chromosome 5 (chr5: 5729959, chr5: 2336903) and two SNPs on chromosome 6 (chr6:1957732, chr6: 3659811) were found to be associated with all three growth traits. Overall, 18 and 17 protein-coding genes were identified for growth traits under the Spanish and Greek environments, respectively. For organ weights, we found a total of 15 SNPs above the suggestive threshold in Spain. Four SNPs (chr22: 1966940, chr22: 6094383, chr22: 67838, chr15: 16798259) were found to be shared between viscera weight and liver weight, while only one SNP (chr22: 67838) were shared between viscera weight and heart weight. None of the SNPs associated with viscera weight, liver weight and heart weight were shared in Greece. Interestingly, the same SNPs were detected to be associated with organ weights and growth traits in the same environment. GO and KEGG functional enrichment analyses showed that light absorption and ECM-receptor interaction are significantly enriched with growth traits QTLs in Spain but not in Greece. For organ weights, response to stimulus is the most prominent process in Spain, while cell adhesion, and regulation of phosphorylation are more prominent in Greece.

Conclusion: Clearly shared genomic architecture across growth traits and organ weights were observed within the Spanish and Greek farms respectively, but we identified totally different genomic architectures for growth traits and organ weights between farms. The involvement of different pathways in the Spanish farm compared to the Greek farm, suggests mechanisms that can explain the genotype-by-environment interaction. These findings will not only explain part of the G x E

interactions, but also give a genetic insight on how differences in environmental conditions affect overall growth and organ weights.

Keywords: seabream, growth, organ weights, GWAS, G x E interactions

5.1 Introduction

Gilthead seabream is one of the domesticated fish species of the Sparidae family. The domestication of gilthead seabream started in the 1970's based on larvae and juveniles capture from the wild (Janssen et al., 2017). Selective breeding, however, was not initiated until the middle of the 1990's (Knibb et al., 1997). Gilthead seabream has developed into one of the main aquaculture finfish species in Europe, with a total volume of 228.000 tons in 2018 (FAO., 2020), ranking fourth in European aquaculture production. Although the species is farmed over a wide geographic area, from the North Atlantic Ocean to the shores of the Red Sea, over 80% of the genetically improved fingerlings are derived from a single country, Greece (Janssen et al., 2015).

G x E interactions for economic traits in aquaculture have been widely reported under different rearing environments. For example, significant G x E interaction has been reported in response to rearing environments under the different environmental factors, such as water salinity (fresh and brackish water in rainbow trout with genetic correlation between 0.19 to 0.48 for harvest weight (*Oncorhynchus mykiss*) (Sae-Lim et al., 2013), seawater and freshwater in Atlantic salmon (*Salmo salar*) with genetic correlation between 0.26 and 0.31 (Gonzalez et al., 2022)), and hypoxic and normoxic environments in Nile tilapia with a genetic correlation 0.81 (Mengistu et al., 2020). Given the wide area in which gilthead seabream are cultured, it is likely that G x E interaction is important in Gilthead seabream aquaculture. An intermediate G x E interactions for harvest weight and growth rate was found for gilthead seabream reared in cage and estuary (Elalfy et al., 2021). A previous study (Gulzari et al., 2022), showed moderate G x E interactions for harvest weight (0.45 ± 0.11), fillet weight (0.49 ± 0.12) and liver weight (0.61 ± 0.11) but a weak G x E interaction for heart weight (0.76 ± 0.11).

QTL identification has been widely applied in a wide range of phenotypes in both livestock and aquaculture species. As growth traits directly affect production and economic benefit, there is much effort to unravel the genomic basis of growth traits. In fish species, the somatotropic axis, (consisting of growth hormone, insulin-like growth factor and their carrier proteins) has been shown to play a key role in regulation of growth in growth (reviewed by De-Santis and Jerry (2007)). So far, there have been several studies identifying QTLs for growth-related traits in fish species, such Atlantic salmon (Gutierrez et al., 2012; Tsai et al., 2015), rainbow trout (Salem et al., 2018; Ali et al., 2020), tilapia (Yoshida and Yáñez, 2021;Yu et al., 2021) and carp (Guo et al., 2022; Huang et al., 2020; Wang et al., 2021). QTL analysis has, so far, received limited attention in seabream. Loukovitis et al. (2011) performed the

first study to understand the genetic basis of body weight and sex determination in gilthead seabream, and further verified their results by increasing the number of informative microsatellite markers (Loukovitis et al., 2012). Although such studies potentially may inform marker assisted breeding, the low marker density limits the applicability. Crucially, the consistency of genomic associations in diverse environments resulting from G x E interaction, has received little attention in gilthead seabream, and other aquaculture species.

Identifying the genomic basis of growth in separate parts of the body, such as, e.g. organ weight, could be beneficial for aquaculture breeding programs. Growth is a composite trait that encompasses and is affected by many other traits. Further investigation of growth that includes, for instance, various organs, provides important information on correlations with physiological conditions. Organ weight is indicative for health, as diseases have been shown to change the weight of internal organs (Kumar et al., 2014). On the other hand, organ weight is strongly corelated with growth traits. For example, harvest weight has exhibited strongly positive genetic correlations with liver weight (0.68), heart weight (0.92) and viscera weight (0.80) in gilthead seabream (Gulzari et al., 2022). So far, only a few QTLs studies associated with organ weight, have been conducted, mostly on mice and chicken (Fawcett et al., 2008; Leamy et al., 2002; Neuschl et al., 2007; Moreira et al., 2019). As a consequence, the genomic basis of organ weight in aquaculture species remains unknown.

In a previous study, moderate G x E interaction was found for growth traits and liver weight of gilthead seabream in two distinct commercial production sites in Spain and Greece (Gulzari et al., 2022). In this study, the same experiment was used to identify QTL for growth-related traits (harvest weight, thermal growth coefficient and fillet weight) and organ weights (viscera weight, liver weight and heart weight) in Gilthead seabream. This study investigates the genomic architecture of G x E interaction, with the inclusion of two different production sites, in Greece and Spain, that were stocked from the same breeding population. Here we extend that study using genomic tools, the main objectives of this study were: (1) to discover QTLs associated with growth and organ weight in Spanish and Greek farms, respectively, (2) to reveal underlying biological processes and pathways related to G x E interactions.

5.2 Methods

5.2.1 Study population

The fish in this study were derived from the selective breeding program at Galaxidi Marine Farm, located in Greece, at the Gulf of Corinth (38° 21’ 06.6’’ N 22°23’18.8’’E).

A total of 33 male and 20 female gilthead seabream were used to mass spawn in a broodstock tank on a single day. The fingerlings were raised as a single group to ensure a consistent common environmental effect. At an average weight of 3 g, one batch of 99,000 juveniles was stocked in a sea cage at the Galaxidi farm and the other batch of 84,605 juveniles was transferred and then stocked at the Cudomar farm located in the southeast of Spain (38°25'12'' N 0°20'51'' W), in the Mediterranean Sea. During the grow-out period, the fish were fed with the same commercial diet at both farms. The feed given per fish was adjusted according to biomass at different stages. The management conditions were kept as similar as possible, and the physical environment is considered the main difference in treatment between the two production farms. Environmental conditions were recorded at the two farms during the grow-out period (water temperature was measured on a daily basis, while dissolved oxygen, salinity, and pH were measured on a monthly basis).

After the grow-out period, production traits and organ weights were collected. This was done after 465 days in Greece, and 500 days in Spain. Prior to sampling, fish received a mortal dose of 0.03 mL/L clove oil in an oxygenated tank. A random sample of fish was taken from the same sea cage. In total, 998 fish in Greece and 945 fish in the Spain were used for data collection. Harvest weight (HW) was measured with a weighting scale accurate to 0.5 g. Fin clips were cut and preserved in 75% ethanol and stored at 4 °C until DNA extraction. In addition, fish were gutted, and viscera weight (VW) was measured (accuracy 0.5 g). Liver weight (LW) and heart weight (HeW) were then separated and weighted by a scale accurate to 0.001 g. The gutted fish were filleted, and the fillet weight (one side) was recorded. The total fillet weight (FW) was calculated by multiplying the one-sided fillet weight by two. Measurements were standardized between the two farms, with the exception of fillet weight, which was skin-off in the Greece and skin-on in Spain. Further details on phenotyping are found in Gulzari et al., (2022). Thermal growth coefficient (TGC) was calculated according to the formula (Mayer et al., 2012):

$$\mathbf{TGC} = \left[\frac{\sqrt[2/3]{\mathbf{harvest\ weight}} - \sqrt[2/3]{\mathbf{stocking\ weight}}}{(\mathbf{T\ x\ t})}\right] \times \mathbf{1000} \qquad (\mathbf{1})$$

where $\boldsymbol{T}$ is the average water temperature, and $\boldsymbol{t}$ is the number of days during grow-out period.

5.2.2 DNA extraction, genotyping and quality control

Genomic DNA was isolated from fin clips by Identigen (Dublin, Ireland). The fish were genotyped using the MedFish array, which contains 30 K SNP markers distributed across the Gilthead seabream reference genome (Peñaloza et al., 2021). Before SNP

filtering, only samples with quality control values >= 0.82 and call rate per samples >=97% were retained before SNP filtering as recommended by the Axiom® Analysis Suite (version 5.1).

Further filtered was done using Plink (Purcell et al., 2007), with the following parameters: SNPs with minor allele frequency (MAF) lower than 5%, missing call rates exceeding 10% were removed. The distribution over the genome and density of SNPs that passed filtering were visualized using the CMplot R package (Yin et al., 2021).

5.2.3 Descriptive statistics, population structure and linkage disequilibrium

Phenotypic statistics (mean value, standard error and t-test) for HW, TGC, FW, VW, LW and HeW of 1901 fish were computed in R version 4.0.2. Prior to doing a genome-wide association study (GWAS), external influence factors such as sampling date were assessed in a linear model using Stepwise Algorithm, as previously described (Yu et al., 2021). Since gilthead seabream is a protandrous fish species, all fish were functional males at the first two years (García Hernández et al., 2020). Therefore, sex was not considered as a factor in this study. After fitting the linear model, observations with residuals greater than 3.5 standard deviations were identified and excluded in the subsequent association analysis.

In order to characterize population stratification, e.g. through family relationships, principal component analysis (PCA) was performed using Plink v1.9 (--pca) option and visualized by the scatterplot3d R package (Ligges and Mächler, 2002). Furthermore, linkage disequilibrium (LD) may limit resolution for the genome-wide association study (Newell et al., 2011). LD decay was estimated in Plink v1.9 with the default option except "--ld-window-r2 0", which means not to filter any SNPs based on squared correlations.

5.2.4 Univariate linear mix model for GWAS

To identify SNPs associated with growth-related traits, a genome-wide association study was performed using linear mixed model with option "-lmm", implemented in Gemma (Zhou and Stephens, 2012):

$$\boldsymbol{y} = \boldsymbol{W\alpha} + \boldsymbol{x\beta} + \boldsymbol{K\mu} + \boldsymbol{\varepsilon} \qquad (2)$$

where $\mathbf{y}$ is an vector of n individuals with phenotypes for each trait (HW, TGC, FW, VW, LW and HeW); $\mathbf{W}$ is an n x 6 matrix of fixed effects (including sampling date and top five PCs to correct population structure); $\boldsymbol{\alpha}$ is a vector of corresponding

coefficients; **x** is a vector of the marker genotypes (0, 1, 2) and **β** is the additive genetic effect of the marker; **K** is the genomic kinship matrix based on standardized matrix method; **μ** is an vector of random animal effects; **ε** is an vector of random residuals. Here, we incorporated the top five PCs accounting for population structure, and the kinship as random effect in the model. Combining PCs and kinship have been shown to perform best in false positive association signal compared to only PCs or kinship by several studies (Hoffman, 2013; Lu et al., 2016; Neves et al., 2012). The estimated p-values for genetic effect are based on the Wald tests. Manhattan and quantile-quantile (Q-Q) plots were visualized by R package qqman (https://cran.r-project.org/web/packages/qqman/). Genomic control λ (also known as inflation factor) was computed in order to assess the presence of population stratification, values of $\lambda_{GC} < 1.05$ are generally considered good (Price et al., 2010).

5.2.5 QTL identification

Due to the overly conservative nature of Bonferroni correction, SimpleM method, based on the effective number of independent tests (Gao et al., 2010), was used to calculate the suggestive (1.71E-04) and genome-wide significance (8.54E-06) thresholds. The SNPs above the suggestive threshold were considered candidate QTLs associated with the corresponding traits. For each QTL, the genetic effect (**β**) of the leading SNPs was used to estimate the proportion of genetic variance explained by this peak per the formula below:

$$\mathbf{Vg}\,\% = \frac{\mathbf{2p(1-p)\beta^2}}{\mathbf{\sigma_a^2}} \times \mathbf{100} \qquad (\mathbf{3})$$

where p is the minor allele frequency of the target SNP, and $\boldsymbol{\sigma_a^2}$ is the total genetic variance.

Candidate genes located within a 50 kb windows around the peak SNPs (50 kb up- and 50 kb down-stream) for each QTLs using SnpEff v5.1 (Cingolani et al., 2012), based on the latest *Sparus aurata* genome reference (GCF_900880675.1). Due to the limited information on the Gilthead seabream genome annotation, we explored additional information by looking homologous genes in Zebrafish (Danio rerio). Therefore, BLASTP searches against zebra fish (*Danio rerio*) proteins, based on genome build GCF_000002035.6_GRCz11, were applied. Best-hits with an expected value less than 1E-*6* were regarded as homologous genes (Yandell et al., 2008; Xu et al., 2016).

5.2.6 Gene ontology (GO) and KEGG enrichment analysis

Based on the candidate genes overlapping GWAS peaks, functional assessment of GWAS results was done by Gene Ontology and KEGG enrichment in Metascape (Zhou et al., 2019) and KOBAS-i (Bu et al., 2021), respectively. Enriched terms with a minimum 2 gene overlap with the input gene list, p-value <0.05 and enrichment factor > 1.5 (the ratio between the observed counts and the counts expected by chance) were considered. The significant GO and KEGG pathways were visualized with the ggplot2 R package (https://github.com/tidyverse/ggplot2).

5.3 Results

5.3.1 SNP informativeness and distribution

In total 1920 fish (979 from the Greek farm and 941 from the Spanish farm) were genotyped for 32,359 SNPs. The SNP Array informativeness was determined by calculating minor allele frequency (MAF). The result showed that most SNPs belong to the common (MAF>=0.3) and intermediate (0.3>MAF>=0.1) minor allele frequency groups at 48.2 % and 40.5%, respectively (Fig 1a). After the quality control using the Axiomon® analysis Suite and Plink, 1901 individuals and 29,111 variants passed quality control and were used for subsequent analyses. The distribution of SNPs in the genome was plotted on the *Sparus aurata* genome reference, showing an even distribution with some clusters of higher density across all chromosomes. (Fig 1b). On average, there is one SNP per 28.63 kb region. The highest number of SNPs (1555) is on chromosome 6, while the lowest number of SNPs (1085) is on chromosome 24.

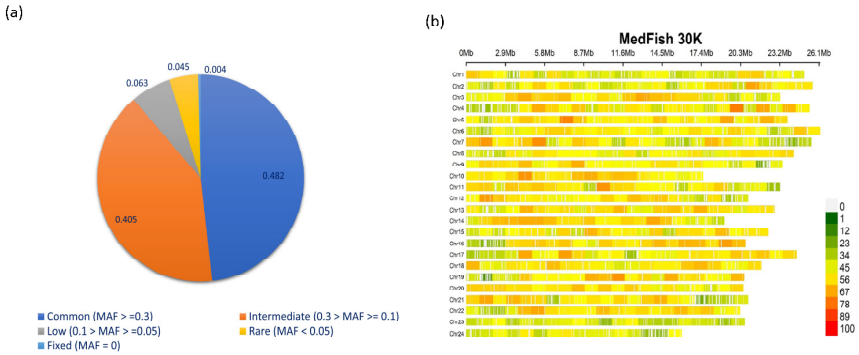


Fig 1. Informativeness (a) and genomic distribution (b) of SNPs used in this study.

5.3.2 Analysis of population structure and LD decay

The top 5 principal components explained 70.04% of the genetic variance of all individuals from both locations combined. According to the PCA plot of the first three

principal components, a mild population stratification was observed (Fig 2a). This bias was corrected by including the top 5 principal components in the GWAS model. Physical linkage of loci can result in linkage disequilibrium (LD). Characterizing LD is useful in GWAS studies as determines the resolution at which loci can be expected to be resolved. The LD decay plot (Fig 2b) shows that average r^2 drops to population background level within 50 kb. Overall, LD is low at the SNP density applied in this study (on average one SNP at every ~30 kb).

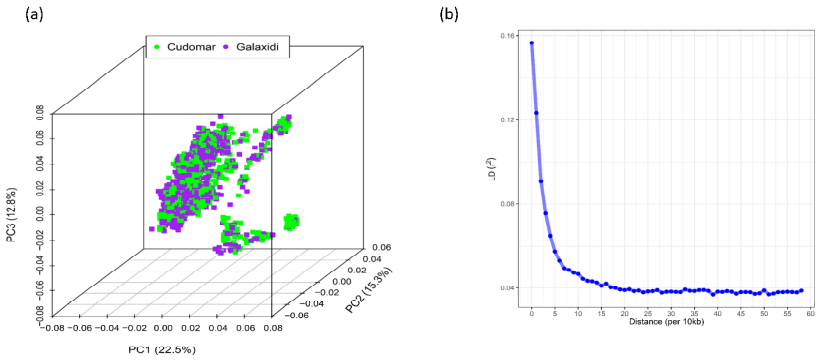


Fig 2. Principle component plot for all individuals across Spanish and Greek farms (a) and estimation of linkage disequilibrium decay using r^2 between pairs of SNPs within 600 kb of each other.

5.3.3 Descriptive statistics

A total of 1901 fish, 934 from Spain and 967 from Greece, passed the quality controls applied for genotypes. For these fish, growth-related traits (HW, TGC, FW) and organ weights (VW, LW, HeW) were measured. As shown in Table 1, fish from the Spanish farm achieved overall mean HW, TGC, FW of 412.31 g, 11.38 ($g^{2/3}$ x $°C^{-1}$ × 1000) and 179.6 g, respectively, while average measurements were 371.44 g, 12.74 ($g^{2/3}$ x $°C^{-1}$ × 1000) and 116.78 g in the Greek farm. Fillet weight showed the highest difference size between two farms. For internal organ weights, liver weight showed the largest difference between two farm locations, followed by heart height. Overall, all six traits were significantly different between the two locations.

Table 1 Descriptive statistics of study traits in two farms

Trait	farm	total number	outliers	mean	se	t-value	effect of farm (P value)
Harvest weight (g)	CUD	934	0	412.31	2.29	13.15	< 2.2e-16
	GMF	967	1	371.44	2.1		
TGC ($g^{2/3}$ x $°C^{-1}$ × 1000)	CUD	934	13	11.38	0.04	-20.06	< 2.2e-16
	GMF	967	9	12.74	0.05		
Fillet weight (g)	CUD	934	3	179.6	0.57	46.10	< 2.2e-16
	GMF	967	3	116.78	0.37		
Viscera weight (g)	CUD	934	4	29.31	0.22	6.68	3.214E-11
	GMF	967	2	27.17	0.23		
Liver weight (g)	CUD	934	4	6.78	0.07	37.38	< 2.2e-16
	GMF	967	4	3.88	0.04		
Heart weight (g)	CUD	934	11	0.48	0.003	22.56	< 2.2e-16
	GMF	967	11	0.38	0.003		

*CUD: Spanish farm; GMF: Greek farm; Mean is after removing outliers.

5.3.4 QTLs for growth traits and organ weights

GWAS was applied to detect loci associated with growth and organ weights. A total of fourteen SNPs were found significantly or suggestively associated with growth traits in Spain (Fig 3a). Seven out of fourteen SNPs were located on chromosome 22, of which the highest peak, which included 3 SNPs between 19.66 Mb to 46.84 Mb, were shared amongst TGC, HW and FW. The leading SNP of the QTLs on chr22 explained an estimated 6.2%, 5.3% and 6.9% of the additive genetic variation for HW, TGC and FW respectively. Four candidate SNPs (chr22: 4686464, chr22: 3468915, chr22: 1966940 and chr11: 10245719) were shared amongst HW, TGC and FW traits. In addition, genomic control factor (λ_{GC}) was close to 1 (0.981, 0.990, 1.014 for HW, TGC and FW), which suggest population stratification is not affecting the association results.

In the Greek farm, a total of 14 SNPs were associated at suggestive level with growth-related traits (Fig 3b). However, none of those 14 SNPs were the same compared to those found for the Spanish farm. Two suggestive peaks, included two SNPs on chromosome 5 (chr5: 5729959, chr5: 2336903) and two SNPs on chromosome 6 (chr6:1957732, chr6: 3659811), were found to be associated with all three traits. The leading SNPs on chromosome 5 explained 8.69%, 8.05% and 10.05% of the additive genetic variation for HW, TGC and FW. Furthermore, eight out of fourteen SNPs are located on chromosome 5 and 6. Similarly as for the Spanish farm, the inflation factors were 1.004, 1.005 and 0.995 for HW, TGC and FW respectively, which suggest the absence of population stratification.

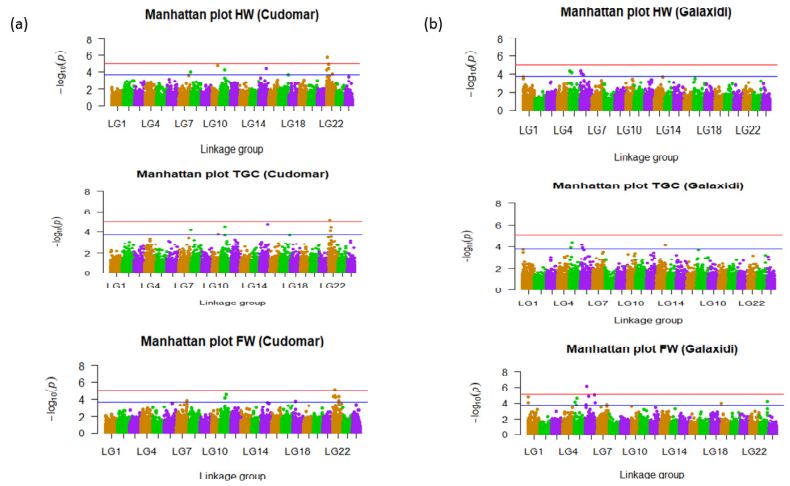


Fig 3. (a) Manhattan plots for HW, TGC, FW for Spain farm. (b) Manhattan plots for HW, TGC, FW for the Greece farm. The orange and blue horizontal line represent the genome-wide significance (5.07) and suggestive threshold value (3.77), respectively.

For organ weights, we found a total of 15 and 17 SNPs above the suggestive threshold in Spain (Fig 4a) and Greece (Fig 5b), respectively. However, similar to the growth-related traits, none of these SNPs were shared between the farms. In Spain, the leading SNPs associated with viscera weight and liver weight are on the same chromosome, but different location (chr24: 7543571 and chr24: 2799809). Four SNPs (chr22: 1966940, chr22: 6094383, chr22: 67838, chr15: 16798259) are shared between viscera weight and liver weight, while only one SNP (chr22: 67838) is shared between viscera weight and heart weight. In Greece, none of the SNPs associated with viscera weight, liver weight and heart weight are shared between traits. Additionally, fewer candidate SNPs were found in the association analyses for Greece compared to Spain, which either reflects less genetic variation, or that the traits are more polygenic in Greece.

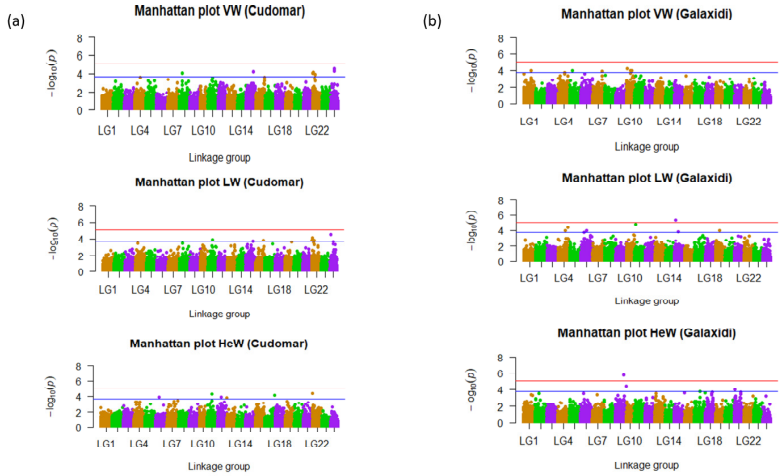


Fig 4. (a) Manhattan plots for viscera weight (VW), liver weight (LW), heart weight (HeW) for Spanish farm. (b) Manhattan plots for VW, LW, HeW for Greek farm. The orange and blue horizontal line represent the genome-wide significance (5.07) and suggestive threshold value (3.77), respectively.

5.3.5 Candidate genes in the QTL regions

A summary of all candidate genes in those QTL regions and their functions is shown in Table 2. In total, eighteen genes were found for those 14 SNPs that were significantly or suggestively associated with growth traits in the Spanish farm. The highest peak was located in an intronic region of *znf292* on chr22 for HW and TGC, and for FW at an intronic region of *lama4* on chr22. Other candidate genes on chr22 are *usp18*, *mcl1b*, *si*, *mapre3b*. Candidate genes on chr11 are *ncanb* and *mob3c*, while the remaining candidate genes are located on other chromosomes. The functional annotation for these candidate genes suggests involvement of laminin complex, tissue development, brain morphogenesis, and apoptotic signaling pathways.

Table 2 Candidate genes for growth traits in the Spain

chr	rs	ps	candidate genes	protein production description
7	AX-325704272	21263574	Intron (*kaznb*)	kazrin, periplakin interacting protein b
8	AX-383841266	819	Intergenic (chr_start; *lamb1a*)	laminin, beta 1a
10	AX-325829954	12200425	Intron (*LOC115589357*)	Immunoglobulin-like fold
11	AX-325864416	10245719	Intron (*ncanb*)	neurocan b
11	AX-384019726	13046556	5_prime_UTR (*mob3c*)	MOB kinase activator 3C
15	AX-384213348	16798259	Intergenic (*LOC562831*; *cfap46*)	gamma-aminobutyric acid receptor subunit pi-like; cilia and flagella associated protein 46
18	AX-326147214	4995960	Intron (*glra1*)	glycine receptor, alpha 1
22	AX-384467133	1966940	Intron (*znf292b*)	zinc finger protein 292b
22	AX-326257091	4686464	Intron (*lama4*)	laminin, alpha 4
22	AX-326255828	3468915	Intergenic (*LOC115574565*; *trnav-cac-5*)	noncoding RNA; noncoding RNA
22	AX-326250505	67838	Intergenic (*usp18*; *mcl1b*)	ubiquitin specific peptidase 18; MCL1 apoptosis regulator, BCL2 family member b
22	AX-384487970	11920943	Intergenic (*si*; *LOC115574686*)	*sucrase-isomaltase*; noncoding RNA
22	AX-326258252	5128085	Exon (trnad-guc-25)	noncoding RNA
22	AX-326271807	12048260	Intron (mapre3b)	microtubule-associated protein, RP/EB family, member 3b

In the Greek farm, sixteen genes were found close to those 14 SNPs that are above the suggestive level for growth traits (Table 3). Candidate locations were found mostly on chr5 and 6. Two peaks were completely overlapping between HW and TGC. Genes located near these peaks are *traf2a*, *tnc, zfr* on chr5 and *uqcc1*, *gdf5* and two non-coding genes (LOC115583312, LOC115583301) on ch6. Other candidate genes are *hmgcra* on chr 5, *gnai2b* and *si* on chr6, while the remainder of genes were found on other chromosomes.

For organ weights, the Spanish farm revealed sixteen candidate genes near the fifteen associated SNPs (as shown in Supplementary Table 1). The only SNP shared between VW, LW, and HeW (chr22: 67838) was located in an intergenic region, between genes *usp18* and *mcl1b*. For the Greek farm, twenty candidate genes were identified for three traits (VW, LW, HeW) (Supplementary Table 2). No shared SNPs were found for these three traits. Interestingly, candidate SNPs were found to be overlapped for growth traits and organ weights, including *ncan* (chr11:10245719), *znf292* (chr22:1966940) and *zfr* (chr5: 5729959) suggesting a functional relationship between growth and organ weight.

Table 3. Candidate genes above suggestive association with growth traits in Greece.

chr	rs	ps	candidate genes	Protein product description
1	AX-325365571	931499	Intron (*LOC559196*)	NACHT, LRR and PYD domains-containing protein 12-like
1	AX-325365525	940207	Intron (*LOC559196*)	NACHT, LRR and PYD domains-containing protein 12-like
5	AX-383652807	2336903	Intergenic (*traf2a; tnc*)	Tnf receptor-associated factor 2a; tenascin Cb
5	AX-325573574	4200955	Exon (*hmgcra*)	3-hydroxy-3-methylglutaryl-CoA reductase a
5	AX-383658435	5729959	Intron (*zfr*)	zinc finger RNA binding protein
6	AX-325618171	1957732	Intergenic (*uqcc1*; *gdf5*)	ubiquinol-cytochrome c reductase complex assembly factor 1; growth differentiation factor 5
6	AX-383716149	3659811	noncoding	-
6	AX-383722376	7054131	noncoding	-
6	AX-383767540	19911726	Intron (*si*)	sucrase-isomaltase
6	AX-383767025	21248353	Intron (*gnai2b*)	guanine nucleotide binding protein (G protein), alpha inhibiting activity polypeptide 2b
7	AX-383834295	20411653	Intron (*cntn3a.2*)	contactin 3a, tandem duplicate 2
13	AX-384125859	18533562	Intergenic (*msantd2 - robo3*)	Myb/SANT DNA binding domain containing 2; roundabout, axon guidance receptor, homolog 3
19	AX-383049895	495763	5'_primer UTR (*LOC110438316*)	zinc finger MYM-type protein 1-like
23	AX-326325158	17713731	Exon (*zgc:136410*)	rhamnose binding lectin-like precursor

5.3.6 GO and KEGG enrichment analyses

Since the number of genes associated with any of the traits was limited, all candidate genes associated growth traits and organ weights were explored to reveal functional enrichments. A total of 59 genes were included in the gene set enrichment. Overall, twelve GO terms were significantly enriched including cell adhesion (GO:0007155, $P < 0.001$), light absorption (GO:0016037, $P< 0.001$), muscle attachment (GO:0016203, $P= 0.007$), and regulation of phosphorylation (GO:0042325, $P= 0.014$) (Fig 5a, Supplementary Table 3). Based on the KEGG pathway database, eight KEGG pathways were significantly enriched (Fig 5b, Supplementary Table 4), including ECM-receptor interaction (dre04512, $P< 0.001$), progesterone-mediated oocyte maturation (dre04914, $P= 0.011$).

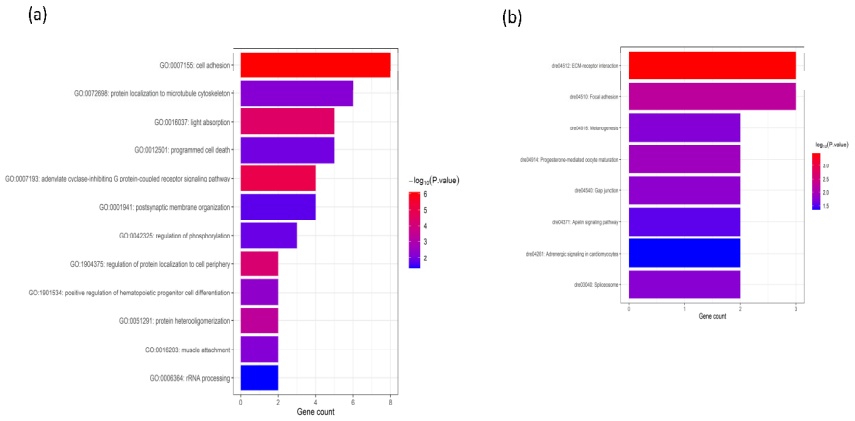


Fig 5. GO (a) and KEGG (b) significant enrichment results by using all candidate genes for growth-related traits and organ weights.

5.4 Discussion

The consistent differences in culture environment between Spain and Greece, provides a relevant experiment to investigate G x E interaction. In this study, two commercial production locations were originally chosen to analyze the effect of divergent temperature effect on growth traits and organ weight, however, we also found other environmental factors such as dissolved oxygen and salinity were significantly different between the two locations. Since the fish were derived from the same broodstock, this provided the opportunity to investigate G x E interaction.

Historical temperature data shows that marine environments of Greece are warmer than the Mediterranean coast of Spain, which should make Greece a more favorable environment for achieving fast growth (Besson et al., 2016). Couto et al. (2008) showed that growth rate of gilthead seabream was higher when reared at 25 rather than at 18 degrees centigrade. However, the temperature at the Greek farm was 0.9 ~1.7 degree lower than that of the Spanish farm during the experiment. This is due to the Greek farm being located at the Gulf of Corinth, where cold water flows down from the mountains in winter. Therefore, a consistently higher average daily water temperatures were found at the Spanish farm compared to the Greek farm, which lead to a significant difference in cumulative degree days. Furthermore, the dissolved oxygen at the Spanish location was also higher than that of the Greek farm. And the measured salinity level at the Greek farm is on average 1.7 ‰ higher than that at the Spanish farm, which is in line with a previous study (Soukissian et al., 2017).

Our results highlight several genes associated with growth traits under two different temperature regimes. For the Spanish farm, our results have shown that a significant

peak was found on chr22 containing genes *znf292* and *lama4*, while four genes *usp18*, *mcl1b*, *si* and *mapre3b* are also prominent in the Spanish farm. Gene *znf292,* was found to be an enhancer of growth hormone expression in rat (Lipkin et al., 1993). Gene *lama4* is expressed in the basement membrane of skeletal muscles (Knöll et al., 2007). For the Greek farm, our results indicate that two suggestive peaks located on chr5 and chr6, respectively. Genes located near these peaks are *traf2a*, *tnc*, *zfr* on chr5, and *uqcc1*, *gdf5* on chr6. Gene *traf2a,* is from *traf* gene family, which has been shown to participate in the immune system and apoptotic process (Nie et al., 2022). Tenascin-C has been shown to regulate the activity of growth factors, such as epidermal growth factor (EGF)-dependent cell growth (Bradshaw, 2014). Gene *gdf5* can be involved in skeletal system, while *zfr* can be involved in multicellular organism development annotated by ZFIN database. Additionally, gene *si,* which encodes sucrase-isomaltase, is found to be important candidate gene for both locations. Pascon et al. (2021) showed that there is a relation between digestion and sucrase-isomaltase activity in European sea bass. Generally, these genes are associated with growth and development under different temperatures, rather than a response to temperature/heat stress.

Different processes and pathways were found to be prominent for growth traits for the Spanish and Greek farms, respectively. In Spain, the candidate genes are significantly enriched for cell adhesion, light absorption, and ECM-receptor interaction. Light (in intensity, quality and photoperiod) is an interesting characteristics in the aquaculture environment and was showed to affect fish growth rate (Sumpter, 1992). Cell adhesion is known to interact with growth factor receptors (Ivaska and Heino, 2010). Interestingly, extracellular matrix (ECM) receptor is known to play an important role in several stress response. For instance, Asakawa et al.,(2019) reported that ECM-receptor interaction was significantly activated after heat stress treatment in Yamame (*Oncorhynchus masou*). Another study show that Extracellular matrix (ECM)-receptor was a enriched pathways response to salinity stress in a hybrid tilapia (Su et al., 2020). In Greece, focal adhesion, muscle attachment and regulation of phosphorylation were significantly enriched in QTL regions for growth-related traits. In general, GO terms and KEGG pathways related to development and growth were commonly identified in both locations. Intriguingly, pathways involved in light absorption and ECM-receptor interaction indicate there may be differences between the farm not related to temperature which were not recorded.

For organ weights, our results indicate several candidate genes involved in organ growth and development, and response to stimulus for the Spanish farm. Gene

ddx49, with highest peak for liver weight, is involved in rRNA processing. Recently, also found that *ddx49* is hub gene for muscle growth in Indian major carp (Mohindra et al., 2022). Gene *znf292b* was shown to be a growth hormone-dependent transcription factor(Lee et al., 2016). Gene *polrmt,* associated with liver weight in our study, has been reported to regulate mitochondrial and energy metabolism (H. J. Yu et al., 2021). Gene *usp18,* which encodes ubiquitin-specific peptidase 18, has been shown to have several biological functions during cell and organ development (Honke et al., 2016). The gene *igsf21* codes for immunoglobin superfamily, member 21a. A GWAS study by Wan et al. (2019) showed that *igsf21* is a candidate gene of disease resistance in large yellow croaker. Interestingly, the same gene *znf292b* is found to regulate both organ weight and growth traits in Spain. The overlap between growth traits and organ weights within locations is expected, as bigger fish have relatively large organs. These results indicated that organ weights are regulated by growth and development pathway coinciding with general growth traits. Furthermore, a warmer temperature may not reach the level of severe stress for the fish, pathways involved in environmental stimulus nevertheless seem to be involved. Although presumably the stressor may be temperature, a direct relationship cannot be inferred.

By contrast, in Greece, twenty genes associated with organ weight were found to be involved in organ growth and development, cell adhesion and regulation of phosphorylation. Genes *birc6* overlapped the significant peak associated with liver weight. Gene *birc6* is known to be involved in apoptotic and ubiquitin-dependent protein catabolic process based on the annotation of ZFIN database. Gene *zfr* involved in multicellular organism development annotated by ZFIN database. The gene *erbb4b* play a critical role in cardiac development, cardiomyocyte proliferation, and homeostasis and function of heart in human (Wadugu and Kühn, 2012). Gene *si* plays an important role in organ weight as well. Considering the genetic correlation between growth and organ weight estimated by Gulzari et al. (2022), a shared genomic architecture between growth and organ weight traits is likely. Interestingly, we did not observe any process related to response to stimulus or stress in Greece, indicating gilthead seabream grow under a relatively comfortable condition.

Overall, difference in enriched pathway for growth traits and organ weight are observed between the two farms. For growth traits, the pathways such as light absorption and ECM-receptor interaction are important in Spain, where temperature is warmer. Different pathways, such as muscle attachment and regulation of phosphorylation, are more prominent in Greece. For organ weights, organ growth and development are both found in two farms. But response to

stimulus was found to be more important in Spain, and cell adhesion, and regulation of phosphorylation in Greece. The different pathways enriched by candidate genes associated with growth traits and organ weights in two farms, suggest some degree of G x E that results in re-ranking of genotypes between environments.

5.5 Conclusions

Although shared genomic architecture for growth traits and organ weights is evident between the two sites, there are different genes involved that may explain the G x E interactions. For growth-related traits, the functional enrichment confirms that light absorption and ECM-receptor interaction processes are important in Spain, with muscle attachment and regulation of phosphorylation in Greece. For organ weight traits, we find organ growth and development processes involved in both farms, while response to stimulus is prominent in Spain, and cell adhesion and regulation of phosphorylation in Greece. Overall, these findings will not only explain part of the G x E interaction, but also give a genetic insight on how environmental differences affect overall growth and organ weights. The QTLs identified in this study provide an insight on genomic basis of fish growth traits and organ weight under a challenging environment.

5.6 Funding

This project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No 727315. Xiaofei Yu has received the financial support from China Scholarship Council (CSC, File No. 201803250069).

5.7 Declaration of Competing Interest

Wageningen University & Research advises Galaxidi Marine Farm on their gilthead seabream breeding program.

5.8 Acknowledgements

The authors thank to the Galaxidi Marine and Cudomar Farms for providing help and facilities for raising the Gilthead seabream and the experimental sampling.

5.9 Supplementary Information

For a compact layout, I have not included all Supplementary material in the thesis, only those which may assist the reader. For more information, the complete Supplementary figures and tables of the article are available at the Open Science Framework repository: https://osf.io/s5yt6/.

Supplementary Table 1. Candidate genes above suggestive association with organ weights in Spain.

Supplementary Table 2. Candidate genes above suggestive association with organ weights in Greece.

Supplementary Table 3. GO significant enrichment with all genes.

Supplementary Table 4. KEGG significant enrichment with all genes.

References

Ali, A., Al-Tobasei, R., Lourenco, D., Leeds, T., Kenney, B., Salem, M., 2020. Genome-wide identification of loci associated with growth in rainbow trout. BMC Genomics 21, 1–16. https://doi.org/10.1186/s12864-020-6617-x

Asakawa, S., Kraitavin, W., Yoshitake, K., Igarashi, Y., Mitsuyama, S., Kinoshita, S., Kambayashi, D., Watabe, S., 2019. Transcriptome analysis of yamame (Oncorhynchus masou) in normal conditions after heat stress. Biology 8, 1–15. https://doi.org/10.3390/biology8020021

Besson, M., Vandeputte, M., van Arendonk, J.A.M., Aubin, J., de Boer, I.J.M., Quillet, E., Komen, H., 2016. Influence of water temperature on the economic value of growth rate in fish farming: The case of sea bass (Dicentrarchus labrax) cage farming in the Mediterranean. Aquaculture 462, 47–55. https://doi.org/10.1016/j.aquaculture.2016.04.030

Bradshaw, A.D., 2014. Regulation of cell behavior by extracellular proteins, in: Principles of Tissue Engineering. Elsevier, pp. 279–290.

Bu, D., Luo, H., Huo, P., Wang, Z., Zhang, S., He, Z., Wu, Y., Zhao, L., Liu, J., Guo, J., 2021. KOBAS-i: intelligent prioritization and exploratory visualization of biological functions for gene enrichment analysis. Nucleic acids research 49, W317–W325.

Cingolani, P., Platts, A., Wang, L.L., Coon, M., Nguyen, T., Wang, L., Land, S.J., Lu, X., Ruden, D.M., 2012. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3. Fly 6, 80–92.

Couto, A., Enes, P., Peres, H., Oliva-Teles, A., 2008. Effect of water temperature and dietary starch on growth and metabolic utilization of diets in gilthead sea bream (Sparus aurata) juveniles. Comparative Biochemistry and Physiology - A Molecular and Integrative Physiology 151, 45–50. https://doi.org/10.1016/j.cbpa.2008.05.013

De-Santis, C., Jerry, D.R., 2007. Candidate growth genes in finfish - Where should we be looking? Aquaculture 272, 22–38. https://doi.org/10.1016/j.aquaculture.2007.08.036

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. https://doi.org/10.1016/j.aquaculture.2021.736462

FAO., A., 2020. The State of World Fisheries and Aquaculture 2020 (Sustainability in Action). Rome.

Fawcett, G.L., Roseman, C.C., Jarvis, J.P., Wang, B., Wolf, J.B., Cheverud, J.M., 2008. Genetic architecture of adiposity and organ weight using combined generation QTL analysis. Obesity 16, 1861–1868. https://doi.org/10.1038/oby.2008.300

Gao, X., Becker, L.C., Becker, D.M., Starmer, J.D., Province, M.A., 2010. Avoiding the high Bonferroni penalty in genome-wide association studies. Genetic Epidemiology: The Official Publication of the International Genetic Epidemiology Society 34, 100–105.

García Hernández, M.P., Cabas, I., Rodenas, M.C., Arizcun, M., Chaves-Pozo, E., Power, D.M., García Ayala, A., 2020. 17α-ethynylestradiol prevents the natural male-to-female sex change in gilthead seabream (Sparus aurata L.). Scientific Reports 10, 1–15. https://doi.org/10.1038/s41598-020-76902-9

Gonzalez, C., Gallardo-Hidalgo, J., Yáñez, J.M., 2022. Genotype-by-environment interaction for growth in seawater and freshwater in Atlantic salmon (Salmo salar). Aquaculture 548, 1–5. https://doi.org/10.1016/j.aquaculture.2021.737674

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

Guo, J., Wang, A., Mao, S., Xu, X., Li, J., Shen, Y., 2022. Construction of high-density genetic linkage map and QTL mapping for growth performance in black carp (Mylopharyngodon piceus). Aquaculture 549, 737799. https://doi.org/10.1016/j.aquaculture.2021.737799

Gutierrez, A.P., Lubieniecki, K.P., Davidson, E.A., Lien, S., Kent, M.P., Fukui, S., Withler, R.E., Swift, B., Davidson, W.S., 2012. Genetic mapping of quantitative trait loci (QTL) for body-weight in Atlantic salmon (Salmo salar) using a 6.5K SNP array. Aquaculture 358–359, 61–70. https://doi.org/10.1016/j.aquaculture.2012.06.017

Hoffman, G.E., 2013. Correcting for Population Structure and Kinship Using the Linear Mixed Model: Theory and Extensions. PLoS ONE 8. https://doi.org/10.1371/journal.pone.0075707

Honke, N., Shaabani, N., Zhang, D.E., Hardt, C., Lang, K.S., 2016. Multiple functions of USP18. Cell Death and Disease 7, 1–9. https://doi.org/10.1038/cddis.2016.326

Huang, X., Jiang, Y., Zhang, W., Cheng, Y., Wang, Y., Ma, X., Duan, Y., Xia, L., Chen, Y., Wu, N., Shi, M., Xia, X.Q., 2020. Construction of a high-density genetic map and mapping of growth related QTLs in the grass carp (Ctenopharyngodon idellus). BMC Genomics 21, 1–12. https://doi.org/10.1186/s12864-020-6730-x

Ivaska, J., Heino, J., 2010. Interplay between cell adhesion and growth factor receptors: From the plasma membrane to the endosomes. Cell and Tissue Research 339, 111–120. https://doi.org/10.1007/s00441-009-0857-z

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. Citeseer.

Knibb, W., Gorshkova, G., Gorshkov, S., 1997. Selection for growth in the gilthead seabream, Sparus auratai L. Israeli Journal of Aquaculture 49, 57–66.

Knöll, R., Postel, R., Wang, J., Krätzner, R., Hennecke, G., Vacaru, A.M., Vakeel, P., Schubert, C., Murthy, K., Rana, B.K., 2007. Laminin-α4 and integrin-linked kinase mutations cause human cardiomyopathy via simultaneous defects in cardiomyocytes and endothelial cells. Circulation 116, 515–525.

Kumar, V., Abbas, A.K., Fausto, N., Aster, J.C., 2014. Robbins and Cotran pathologic basis of disease, professional edition e-book. Elsevier health sciences.

Leamy, L.J., Pomp, D., Eisen, E.J., Cheverud, J.M., 2002. Pleiotropy of quantitative trait loci for organ weights and limb bone lengths in mice. Physiological Genomics 2002, 21–29. https://doi.org/10.1152/physiolgenomics.00018.2002

Lee, J.H., Song, S.Y., Kim, M.S., Yoo, N.J., Lee, S.H., 2016. Frameshift mutations of a tumor suppressor gene ZNF292 in gastric and colorectal cancers with high microsatellite instability. Apmis 124, 556–560. https://doi.org/10.1111/apm.12545

Ligges, U., Mächler, M., 2002. Scatterplot3d-an r package for visualizing multivariate data. Technical Report.

Lipkin, S.M., Näär, A.M., Kalla, K.A., Sack, R.A., Rosenfeld, M.G., 1993. Identification of a novel zinc finger protein binding a conserved element critical for Pit-1-dependent growth hormone gene expression. Genes & development 7, 1674–1687.

Loukovitis, D., Sarropoulou, E., Batargias, C., Apostolidis, A.P., Kotoulas, G., Tsigenopoulos, C.S., Chatziplis, D., 2012. Quantitative trait loci for body growth and sex determination

in the hermaphrodite teleost fish Sparus aurata L. Animal Genetics 43, 753–759. https://doi.org/10.1111/j.1365-2052.2012.02346.x

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. https://doi.org/10.1371/journal.pone.0016599

Lu, K., Xiao, Z., Jian, H., Peng, L., Qu, C., Fu, M., He, B., Tie, L., Liang, Y., Xu, X., Li, J., 2016. A combination of genome-wide association and transcriptome analysis reveals candidate genes controlling harvest index-related traits in Brassica napus. Scientific Reports 6, 1–13. https://doi.org/10.1038/srep36452

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

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/https://doi.org/10.1016/j.aquaculture.2020.735704

Mohindra, V., Chowdhury, L.M., Chauhan, N., Maurya, R.K., Jena, J.K., 2022. Transcriptome analysis revealed hub genes for muscle growth in Indian major carp, Catla catla (Hamilton, 1822). Genomics 114, 110393.

Moreira, G.C.M., Salvian, M., Boschiero, C., Cesar, A.S.M., Reecy, J.M., Godoy, T.F., Ledur, M.C., Garrick, D., Mourão, G.B., Coutinho, L.L., 2019. Genome-wide association scan for QTL and their positional candidate genes associated with internal organ traits in chickens. BMC Genomics 20, 1–15. https://doi.org/10.1186/s12864-019-6040-3

Neuschl, C., Brockmann, G.A., Knott, S.A., 2007. Multiple-trait QTL mapping for body and organ weights in a cross between NMRI8 and DBA/2 mice. Genetical Research 89, 47–59. https://doi.org/10.1017/S001667230700852X

Neves, H.H.R., Carvalheiro, R., Queiroz, S.A., 2012. A comparison of statistical methods for genomic selection in a mice population. BMC Genetics 13, 1–24. https://doi.org/10.1186/1471-2156-13-100

Newell, M.A., Cook, D., Tinker, N.A., Jannink, J.-L., 2011. Population structure and linkage disequilibrium in oat (Avena sativa L.): implications for genome-wide association studies. Theoretical and Applied Genetics 122, 623–632.

Nie, Z., Wang, B., Zhang, Z., Jia, Z., Xu, R., Wang, H., Zhou, W., Gong, Y., 2022. Genome-wide identification of the traf gene family in yellow catfish (Pelteobagrus fulvidraco) and analysis of their expression in response to bacterial challenge . Journal of Fish Biology 0–1. https://doi.org/10.1111/jfb.15126

Pascon, G., Messina, M., Petit, L., Valente, L.M.P., Oliveira, B., Przybyla, C., Dutto, G., Tulli, F., 2021. Potential application and beneficial effects of a marine microalgal biomass produced in a high-rate algal pond (HRAP) in diets of European sea bass, Dicentrarchus labrax. Environmental Science and Pollution Research 28, 62185–62199. https://doi.org/10.1007/s11356-021-14927-x

Peñaloza, C., Manousaki, T., Franch, R., Tsakogiannis, A., Sonesson, A.K., 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 113, 2096–2107. https://doi.org/10.1016/j.ygeno.2021.04.038

Price, A.L., Zaitlen, N.A., Reich, D., Patterson, N., 2010. New approaches to population stratification in genome-wide association studies. Nature Reviews Genetics 11, 459–

463. https://doi.org/10.1038/nrg2813

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A.R., Bender, D., Maller, J., Sklar, P., De Bakker, P.I.W., Daly, M.J., 2007. PLINK: a tool set for whole-genome association and population-based linkage analyses. The American journal of human genetics 81, 559–575.

Sae-Lim, P., Kause, A., Mulder, H.A., Martin, K.E., Barfoot, A.J., Parsons, J.E., Davidson, J., Rexroad, C.E., Van Arendonk, J.A.M., Komen, H., 2013. Genotype-by-environment interaction of growth traits in rainbow trout (Oncorhynchus mykiss): A continental scale study. Journal of Animal Science 91, 5572–5581. https://doi.org/10.2527/jas.2012-5949

Salem, M., Al-Tobasei, R., Ali, A., Lourenco, D., Gao, G., Palti, Y., Kenney, B., Leeds, T.D., 2018. Genome-wide association analysis with a 50K transcribed gene SNP-chip identifies QTL affecting muscle yield in rainbow trout. Frontiers in Genetics 387.

Soukissian, T.H., Denaxa, D., Karathanasi, F., Prospathopoulos, A., Sarantakos, K., Iona, A., Georgantas, K., Mavrakos, S., 2017. Marine renewable energy in the Mediterranean Sea: Status and perspectives. Energies 10. https://doi.org/10.3390/en10101512

Su, H., Ma, D., Zhu, H., Liu, Z., Gao, F., 2020. Transcriptomic response to three osmotic stresses in gills of hybrid tilapia (Oreochromis mossambicus female × O. urolepis hornorum male). BMC Genomics 21, 1–20. https://doi.org/10.1186/s12864-020-6512-5

Sumpter, J.P., 1992. Control of growth of rainbow trout (Oncorhynchus mykiss). Aquaculture 100, 299–320.

Tsai, H.Y., Hamilton, A., Guy, D.R., Tinch, A.E., Bishop, S.C., Houston, R.D., 2015. The genetic architecture of growth and fillet traits in farmed Atlantic salmon (Salmo salar). BMC Genetics 16. https://doi.org/10.1186/s12863-015-0215-y

Wadugu, B., Kühn, B., 2012. The role of neuregulin/ErbB2/ErbB4 signaling in the heart with special focus on effects on cardiomyocyte proliferation. American Journal of Physiology - Heart and Circulatory Physiology 302, 2139–2147. https://doi.org/10.1152/ajpheart.00063.2012

Wan, L., Wang, W., Liu, G., Dong, L., Li, W., Han, Z., Ye, K., Wang, Z., 2019. A genome-wide association study of resistance to Pseudomonas plecoglossicida infection in the large yellow croaker (Larimichthys crocea). Aquaculture International 27, 1195–1208. https://doi.org/10.1007/s10499-019-00376-4

Wang, L., Jia, S., Zhang, Y., Jiang, S., Chen, Y., Chen, J., Yu, M., Zhang, L., Qiao, Z., Li, X., 2021. Construction of three high-density genetic linkage maps and dynamic qtl mapping of growth traits in yellow river carp (Cyprinus carpio haematopterus). Current Issues in Molecular Biology 43, 2276–2288. https://doi.org/10.3390/cimb43030160

Xu, M., Jiang, L., Shen, K.N., Wu, C., He, G., Hsiao, C. Der, 2016. Transcriptome response to copper heavy metal stress in hard-shelled mussel (Mytilus coruscus). Genomics Data 7, 152–154. https://doi.org/10.1016/j.gdata.2015.12.010

Yandell, M., Moore, B., Salas, F., Mungall, C., MacBride, A., White, C., Reese, M.G., 2008. Genome-wide analysis of human disease alleles reveals that their locations are correlated in paralogous proteins. PLoS Computational Biology 4. https://doi.org/10.1371/journal.pcbi.1000218

Yin, L., Zhang, H., Tang, Z., Xu, J., Yin, D., Zhang, Z., Yuan, X., Zhu, M., Zhao, S., Li, X., 2021. rMVP: a memory-efficient, visualization-enhanced, and parallel-accelerated tool for genome-wide association study. Genomics, proteomics & bioinformatics.

Yoshida, G.M., Yáñez, J.M., 2021. Multi-trait GWAS using imputed high-density genotypes from whole-genome sequencing identifies genes associated with body traits in Nile tilapia. BMC Genomics 22, 1–13. https://doi.org/10.1186/s12864-020-07341-z

Yu, H.J., Xiao, G.L., Zhao, Y.Y., Wang, X.X., Lan, R., 2021. Targeting Mitochondrial Metabolism

and RNA Polymerase POLRMT to Overcome Multidrug Resistance in Cancer. Frontiers in Chemistry 9, 1–7. https://doi.org/10.3389/fchem.2021.775226

Yu, X., Megens, H.J., Mengistu, S.B., Bastiaansen, J.W.M., Mulder, H.A., Benzie, J.A.H., Groenen, M.A.M., Komen, H., 2021. Genome-wide association analysis of adaptation to oxygen stress in Nile tilapia (Oreochromis niloticus). BMC Genomics 22, 1–13. https://doi.org/10.1186/s12864-021-07486-5

Zhou, X., Stephens, M., 2012. Genome-wide efficient mixed-model analysis for association studies. Nature genetics 44, 821–824.

Zhou, Y., Zhou, B., Pache, L., Chang, M., Khodabakhshi, A.H., Tanaseichuk, O., Benner, C., Chanda, S.K., 2019. Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. Nature communications 10, 1–10.

6

General discussion

6.1 Introduction

The traditional method of aquaculture breeding is to select fish that grow fast and large at harvest time under controlled environments. Breeding for better growing fish under a challenging environment is usually not considered. The challenging environment refers to the critical parameters significantly different from usual culturing conditions, such as dissolved oxygen, temperature, or salinity, and which may be difficult to control. However, as a result of climate change, it has become urgent to consider breeding goals that enable fish to grow and adapt well to environmental conditions that can unexpectedly change beyond the optimum range very rapidly. Although it should be possible to achieve this breeding goal by traditional selective breeding, such as mass selection or family-based selection, these methods are usually accompanied by a relatively slow genetic response compared to marker-assisted selection (Jeon, 2002; Wakchaure et al., 2015). Thankfully, the rapid development of genomic resources and genotyping strategies provide us with ever better opportunities to investigate the fundamental mechanisms underlying adaptation. Genomic information can potentially be utilized for accelerating the genetic progress in aquaculture species by means of marker-assisted selection.

In this thesis, I focus on unravelling the genomic architecture underlying adaptation to three major environmental challenges in aquaculture: dissolved oxygen, salinity, and temperature. The results presented in this thesis provide insight into the genomic basis of target traits in fish species that indicate a better adaptation to challenging environments. In section **6.2** I discuss the genomic basis of environmental adaptation using forward and reverse genetics approaches (Bomblies and Peichel, 2022) and propose a way how to combine forward and reverse approaches to study the genetic mechanism of adaptation to a challenging environment in aquaculture. In section **6.3** I discuss genotype-by-environment (G X E) interaction from a genomic perspective (including environment-shared and environment-specific QTLs) to further extend our current knowledge and make suggestions on how these could be used in future experiments. In section **6.4** I discuss the advantages and disadvantages of genomic technologies for QTL identification in aquaculture, from SNP array to whole genome sequencing, as used in this thesis, and present a stepwise selection procedure to come to the right approach.

6.2 Genomic architecture of adaption to a challenging environment in aquaculture

The key to genomic architecture of adaptation to a challenging environment is to identify genetic variants underlying variations of traits. There are two approaches that are commonly applied to study the genomics of adaptation to specific environments. These are referred to as forward (or top-down) and reverse (or bottom-up) genetic approaches. The forward genetic approach can be applied when many genetically different individuals with phenotypic records and genotypes if many genomic markers are available. Candidate variants underlying adaptation to a challenging environment are identified by an association between genotype and phenotype, such as QTL mapping and GWAS (Mackay et al., 2009). The reverse genetic approach, also referred to as genome scan, is used to identify genomic regions with selection signatures in individuals from distinct populations (Bomblies and Peichel, 2022). For this approach, phenotypes from different individuals are not needed.

6.2.1 Forward approach

For the forward approach, two types of experiments are used in aquaculture: the acute challenge test, and the acclimatization experiment. In the following sections, i will compare results from these approaches for critically environmental factors, such as oxygen and temperature.

Oxygen adaptation

Oxygen adaptation can be measured by the oxygen challenge test: the time to loss of equilibrium (LOE), where the fish are subjected to an oxygen concentration decline until they lose dorsal–ventral equilibrium (Regan et al., 2017; Bergstedt et al., 2021). A study by Li et al., (2017) found four genome-wide significant and many suggestive QTLs for LOE in Nile tilapia, and revealed two genes significantly associated with LOE. In catfish, Wang et al.(2017) applied a GWAS to identify QTLs associated with LOE across and within strains. Their results showed that rare QTLs were shared across strains and one to two significant SNPs could be detected within strains. In Atlantic salmon, Lin (2017) also found only two SNPs that were associated with LOE. These results reveal the polygenic nature of adaptation to acute hypoxia, measured as LOE, varying from species to species and from strain to strain.

However, the mean duration of hypoxia in an LOE test usually lasts only a few hours, which is far removed from the situation in practice. Therefore, I applied an acclimatization experiment to study the genomic architecture of adaptation to recurrent hypoxia. In the acclimatization experiment, Nile tilapia were exposed to a

daily fluctuation in oxygen where hypoxia occurred especially at the end of the night, for a total of 217 days. In **Chapter 2** I showed a clear association between SNPs and phenotypic variation during adaptation to a hypoxic environment, including 36 candidate SNPs from well-known hypoxia-regulated genes such as *igf1rb*, *rora*, *efna3* and *aurk*. Compared to the LOE challenge test for oxygen adaptation, I believe that the acclimatization experiment captured the candidate region and genes that are important for tolerance or adaptation to low oxygen. This is because the candidate genes identified in the oxygen challenge test are related to acute response to hypoxia, whereas candidate genes identified in the acclimatization experiment are related to growth affected by hypoxia. Moreover, I found more significant and suggestive SNPs than previous studies using challenge tests in Nile tilapia (Li et al., 2017). The differences in results are partially due to the increased power of GWAS, but also shows that there is a difference in the genomic architecture of traits measured by LOE challenge test and the acclimatization experiment.

Temperature adaptation

Temperature adaptation is an important trait for aquaculture and has become even more relevant due to global warming. So far, several studies focused on the identification of QTLs associated with temperature tolerance by applying acute challenges with temperature as stressor, in which the fish are individually monitored for loss of equilibrium (LOE) over time. In this approach, the environmental temperature is controlled by gradually adding heated water to the aquatic system. In catfish, three significant QTLs, located on different linkage groups, were associated with heat tolerance, measured as LOE. These QTLs explained between 11.3 and 12.1 % of phenotype variance (Jin et al., 2017). In Chinook salmon, Everett and Seeb (2014) identified three QTLs for temperature tolerance by a LOE experiment, and these QTLs accounted for 34 % to 64 % of the phenotypic variance, while one QTL located on chr19 was found to be overlapping with a temperature tolerance QTL in rainbow trout (Jackson et al., 1998). Two significant QTLs associated with temperature tolerance in Arctic charr were also found to be homologous to the QTLs in rainbow trout (Somorjai et al., 2003). On the other hand, temperature tolerance is found to be highly variable with a substantial genetic variation among families of Atlantic salmon (Anttila et al., 2013). These results suggest a wide variation in genetics of tolerance to acute hyperthermia across species, and that important QTLs, and therefore genetic architecture of adaptation, can be conserved between related species or vary within a species.

Compared to the challenge test for temperature tolerance above, I successfully captured the genomic architecture through an acclimatization experiment to

relatively small temperature changes (**Chapter 5**). In the experiment, gilthead seabream was exposed to two different temperature regimes in Greece and Spain for a total of 465 and 500 days, respectively. The genomic architecture associated with growth traits of gilthead seabream was identified in two different temperature farms of gilthead seabream (Spanish and Greek farms). Similar to oxygen adaptation, I did not find any overlap between genes identified in LOE challenge tests for acute temperature stress and the acclimatization experiment. A possible explanation is the large evolutionary difference between gilthead seabream and the *salmonid* family. It is also possible, as discussed above, that the candidate genes identified in the challenge test are only related to acute upper thermal temperature tolerance, whereas the candidate genes identified in the acclimatization experiment are related to growth affected by small differences in temperature. Taken together, the results clearly show that the nature of the temperature stress is quite different from two experimental methods.

To summarize, we successfully identified the genomic regions that are important for adaptation to the challenging environments. The differences in genomic architecture of environmental adaptation between the challenge test and the acclimatization experiment indicate that these are two very different traits.

6.2.2 Reverse approach

The reverse approach has been used to identify the genomic basis of adaptation in populations that have been under selection in different environments. This approach has been widely applied to detect signatures of both natural and artificial selection.

Natural selection

Signatures of natural selection provide an insight into adaptation in populations. For example, opah (*Lampris spp*) is one of few fishes that can elevate body temperatures relative to the living environment. A study by Wang et al., (2022) showed that a large number of genes in the opah genome were under positive selection and significantly involved in basis process for whole-body endothermy. Also, more than 100 independent selective signals were identified between spring- and autumn-spawning populations belonging to Atlantic herring of different geographic origin (Barrio et al., 2016). The fish species *Schizothoracinae* is well-adapted to the harsh environment (including hypoxia and low temperature) on the Tibetan Plateau (Guan et al., 2014). Positive selection signatures were detected, that included 162 genes involved in energy metabolism, ion transport, and immune response including the toll-like receptor signaling pathway (Tong et al., 2017). Another interesting example is adaptation to toxic pollution in killifish. Individuals from geographically separated

populations were sampled, and several key genes involved in the mediating toxicity such as aryl hydrocarbon receptor (*AHR*), were shown to be strongly selected (Reid et al., 2016).

Artificial selection

Signatures of artificial selection provide insight into the genomic architecture of selection. Over the past few decades, traditional selection based on phenotypic performance has greatly improved aquaculture production. For example, introducing wild Atlantic salmon into captivity and subsequent artificial selection for production traits has resulted in phenotypic differences between domestic and wild fish. Using whole genome sequencing, Naval-Sanchez et al. (2020) identified 139 sweep regions with reduced heterozygosity in the farmed salmon compared to the wild populations, indicating selection on brain function and behavior. Another study performed by López et al. (2021) in two different lines of farmed coho salmon showed selection signatures at genes associated with body weight, such as *anapc2*, *alad*, *chp2* and *myn*, and genes (*sec24d* and *robo1*) associated with resistance to *Piscirickettsia salmonis*. By comparing the Abbassa Strain of Nile tilapia (after 11 years of selective breeding) to wild Egyptian Nile River populations, approximately 6.9% of SNPs were identified as outliers across all 22 *O. niloticus* chromosomes resulting from selection (Nayfa et al., 2020). These examples indicate that identifying selection signatures can enhance understanding of the genetic mechanisms underlying phenotypic changes from wild to domesticated populations.

It is possible to detect signatures of artificial selection in populations used in long-term breeding programs, but also populations based on more recent breeding programs. For instance, for a breeding program of Atlantic salmon which started in the 1970s, Gutierrez et al. (2016) reported significant evidence for selection signatures which were distributed across 22 of the 29 chromosomes. A genome scan on turbot showed that strong selection for growth was evident after four generations (Vilas et al., 2015). In **Chapter 4** I successfully detected selection signals for salinity adaptation using animals derived from a four-generation breeding program. My results show that the strongly selected regions were overrepresented for pathways related to the regulation of salinity tolerance, such as MAPK3 activity, potassium ion homeostasis, ATPase activity and response to calcium ion.

Overall, the reverse approach seems to be a far more efficient method of detecting the genomic basis of adaptation to a challenging environment. It avoids the challenges to measure-specific traits and overcomes the problem of not knowing the phenotypic target of a selection. It also allows detecting selection signatures across

a wide range (from hundreds of generations down to only a few generations), from wild to domesticated populations.

6.2.3 Combining forward and reverse approach

Although forward and reverse approaches both have their advantages, they also have their disadvantages. For instance, a major disadvantage of the forward approach is the challenge to study the genetic basis of adaptation without phenotypic information (Barrett and Hoekstra, 2011). Moreover, it is also easy to miss traits that are less obvious yet still relevant for adaptation. For the reverse approach, several problems have been previously reported, including technical biases and other factors such as demography and gene flow (Bomblies and Peichel, 2022). On the other hand, as shown above, several hundreds of genes can be identified by a genome scan, making it difficult to determine the importance based on their function.

In the following schematic diagram, I propose to combine forward and reverse approaches to study the genetic mechanism of adaptation to environmental challenges as shown in **Figure 1**. At **step 1**, farmed fish from selective breeding programs are reared in contrasting challenging environments for at least three generations. In view of the challenges, I would suggest focusing on one major challenge like high temperature and then dive into the genomic basis of adaptation to that specific environment. At **step 2**, genome scans (reverse genetic approaches) can be used to detect selection signatures underlying the divergence in adaptive traits. Phenotype-genotype association (forward genetic approach) can be performed for farmed fish from these generations in challenge-type tests through QTL mapping or GWAS. At **step 3**, population genetic approaches can be used to compare between candidate regions from reverse and forward genetic approaches. The genomic regions are targeted if they are under directional selection. At **step 4**, it is necessary to further investigate the functional information for the target region, assisting further determination of the causal variants. Thus, several key data types (including gene expression by RNA-seq, DNA methylation by whole genome bisulfite sequencing,) can be incorporated to prioritize SNPs in a genomic breeding program (see 6.4), as gene expression and methylation have widely been shown to determine local adaptation in both plant and livestock (Kinoshita and Seki, 2014; Dubin et al., 2015; Sejian et al., 2018; Alakärppä et al., 2018; Del Corvo et al., 2021). At **step 5**, the true casual variants and mutants could be validated by a loss-of-function (such as CRISPR-Cas9 system) experiment. As the traits indicating adaptation mostly polygenic, I recommend only continue this step only when major QTLs and genes are

found. At **step 6**, if true casual variants are successfully detected, and the advantageous allele can be selected based on genotypes of the whole population.

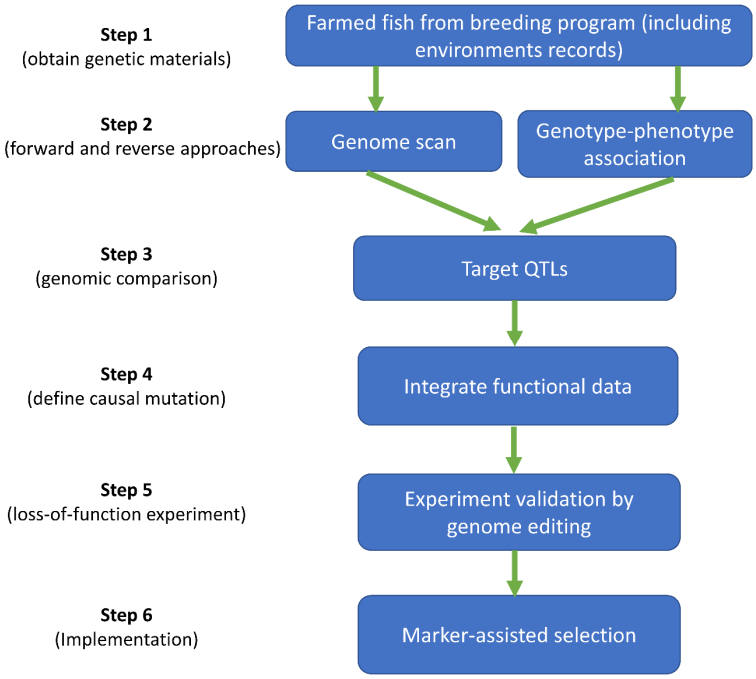


Figure 1. A schematic diagram combing forward and reverse approaches to study and apply the genetic mechanism of adaptation.

Combining the forward and reverse approaches will not only complement signals that cannot be detected by one approach separately, but also provide more confidence with respect to the true QTLs associated with the target traits.

To continuously study the adaptation to the challenging environment in the future, studies need to move from phenotype-genotype association to regulatory network level (gene expression, methylation status). Although we successfully identified several candidate genes during our acclimatization experiments, a lack of functional datasets makes it hard to have a good view of the regulatory network (the key genes that interact with several other genes) that together influence the traits under challenging environments. These key genes in the regulatory network need to be prioritized when applying marker-assisted selection and genomic prediction.

6.3 Genomic view on Genotype by environment interactions

6.3.1 Classic G X E interaction

Genotype by environment (G X E) interaction refers to the difference in the response of genotypes in divergent environments (Falconer, 1996). Two forms of G X E interactions can be distinguished: scaling or re-ranking (Sae-Lim et al., 2016). The re-ranking effect of genotype is the most relevant form of G X E interaction for selective breeding. It means that selection in a single environment will result in lower-than-expected genetic gains and selection response in the other environment. Consequently, a trait can be regarded as a different trait if measured in different environments. From a breeding program perspective, environment-specific programs are recommended if the genetic correlation is lower than 0.7 - 0.8 (Mulder, 2007).

From a quantitative genetic perspective, G X E is well understood for adaptation to different challenging environments, including hypoxia and salinity stress. For instance, the hypoxic environment in non-aerated ponds significantly reduced the harvest weight, survival, and growth rate compared to the normoxic condition in aerated ponds, while the genetic correlation between the two environments was relatively high (0.81 and 0.78) for harvest weight and thermal growth coefficient (Mengistu et al., 2020). However, a moderate G X E interaction was observed for growth traits under salinity stress. Genetic correlations for harvest weight and daily growth coefficient were 0.66 and 0.65 between brackish water and freshwater (Setyawan et al., 2022). More severe G X E interaction was found for environmental temperature. A precious study showed that the genetic correlations for harvest weight and thermal growth coefficient were 0.45 and 0.43 for two different production sites in the Mediterranean region (Gulzari et al., 2022). The genomic mechanisms underlying these G X E interactions are poorly understood.

6.3.2 QTL-by-environment interaction

From a genomic perspective, G x E interaction result from genes having different effects depending on the environment (Lillehammer et al., 2008). There, QTLs can be dissected into three categories based on their genetic by environment effect: environment-specific QTLs, environment-shared QTLs, and neutral QTLs. Ideally, QTLs selected in a breeding program are the environment-shared category because their effect remains stable independent of the environment, allowing the same genetic improvement in a variety of environments.

So far, several studies in plants have reported the identification of QTLs that belong to the environment-shared and environment-specific categories. For example, a in a

cotton inbred line population, a total of 165 QTLs associated with fiber quality traits, of which 47 QTLs were shared across 11 environments, were found (Jamshed et al., 2016). A study by Bao et al., (2018) in rapeseed showed a total of 74 environment-shared and 22 environment-specific QTLs. These results show that QTL can be associated with traits varied from environment to environment, but also can be shared across environments. Consequently, identification of these environment-shared and environment-specific QTLs will help understand on the fundamental mechanism of G X E from biological processes and pathways.

To date, there are no studies on detecting environment-specific and environment-shared QTLs in aquaculture, which is why, in this thesis, genomic architecture of G x E received specific attention. In **Chapter 2,** two groups of the GIFT strain derived from the same families were raised in aerated (normoxic) and non-aerated (hypoxic) ponds. Using a meta-analysis of GWAS results under two environments, we were able to identify environment-shared QTLs associated with the growth of GIFT strain for the two environments. Our results indicated that environment-shared QTLs were found to be over-represented for pathways involved in nervous system development and organ growth in the early stage, and oocyte maturation in the later growth stage. Environment-specific QTLs were over-represented for other pathways, e.g., MAPK and VEGF signalling pathways, and these genes are candidates for adaptation to hypoxia, but are likely not relevant for the normoxic environment. For temperature changes, we also find environment-specific QTLs associated with growth-related traits in Spanish and Greek farms, respectively. The related genes were enriched in light absorption and ECM-receptor interaction in warmer temperature farm (Spanish farm), but not the Greek farm. Although a limited number of QTLs were discovered in this thesis, these results nevertheless provide compelling new insights in the biological mechanism of G x E.

By applying quantitative genetics, G x E can be assessed as genetic correlations. But, by applying genomics, G X E interaction can be revealed by the environment specific SNPs, e.g., in a GWAS analysis. A moderate genetic correlation was reported between two DO environments in **Chapter 2**, while thirty-three SNPs were identified to be environment-shared. Furthermore, shared biological processes/pathways were also observed for two DO environments. On the other hand, a low genetic correlation was found between two different temperature environments, while only one SNP was identified to be environment-shared in **Chapter 5**, without any shared processes/pathways. Although the sample size is relatively small for the experiments described in **Chapter 2** and **Chapter 5**, the number of environmental-specific genes

and the underlying pathways are useful ways to indicate the level of G X E interaction from a genomic perspective.

Overall, these results enlighten the nature of G X E interaction from a genomic perspective, reflected by environment-specific and environment-shared genes. To continue this direction, the environmental factors need to be well recorded across major farms in different countries, especially those that potentially largely impact your target traits. This environmental information provides the opportunity to study in depth for each environment. The effect of environment-specific and environment-shared QTLs on the target traits, as has been shown in the plant breeding (Zaïm et al., 2020).

6.4 Genomic technologies for QTL identification in aquaculture

Genetic improvement in aquaculture has relied on phenotypic and pedigree information over the last few decades, but advanced genomic technologies have gradually been applied to identify genes or markers associated with economically important traits. Currently, the most representative genomic technologies are genotyping-by-sequencing (GBS), SNP arrays and whole genome sequencing (WGS).

GBS has been applied for high throughput, low-cost genotyping. It relies on the fragmentation of the genomic DNA by restriction enzymes to produce a reduced representation of the genome (Elshire et al., 2011). Applications of GBS include the generation of genetic linkage maps, genome-wide association studies, and improvements of reference genome assemblies. More recently it has been applied in aquaculture for genomic selection for traits of interest like growth, sex determination, and disease resistance in a ranges of aquaculture species as summarized by Robledo et al. (2018). However, GBS has several disadvantages. First, although GBS enables the SNPs' identification at a whole genome level, a low genome representation is unavoidable since that is part of the design. Thus, problems such as high genotyping errors, high missing call rates and low accuracy often arise when calling heterozygous SNPs (Wang et al., 2020). This phenomenon is also observed in the results presented in **Chapter 2**. Initially, a total of 95,244 SNPs were generated with 3500 fish. However, after applying all quality control steps, only a minority of those SNPs (27,090) was kept. Second, GBS requires complex bioinformatics analysis. So far, most GBS bioinformatics pipelines such TASSEL-GBS (Glaubitz et al., 2014), GBS-SNP-CROP (Melo et al., 2016), Fast-GBS (Torkamaneh et al., 2017), GB-easy (Wickland et al., 2017), and NGSEP (Perea et al., 2016) originally were designed for application in crops, while no of workflows specifically for

aquaculture species have been designed. A Haplotype-based pipeline was used in **Chapter 2** (Garrison and Marth, 2012). However, this method was reported to detect fewer genetic variants and result in lower coverage compared to other method (e.g. TASSEL-GBS) (Yu et al., 2017).

SNP arrays have undoubtedly become an important tool for large population studies. For instance, commercial SNP arrays have been used for Nile tilapia and gilthead seabream in **Chapter 3** and **Chapter 5**. The SNPs on these arrays are representing all the linkage groups or chromosomes consistently. SNP arrays are regarded as a powerful and high-throughput genotyping technique for several aquaculture species, for which commercial arrays are now available: various salmonids (Houston et al., 2014; Palti et al., 2015; Bernard et al., 2022), gilthead seabream and sea (Penaloza et al., 2021), and Nile tilapia (Joshi et al., 2018; Peñaloza et al., 2020b).

Applications of SNP arrays include constructing linkage maps and studying the genetic architecture of production traits such as growth, disease resistance, and genomic selection and prediction. Compared to GBS, SNP arrays have the advantage of increased genotyping accuracy (Peñaloza et al., 2020a). However, the main drawback of SNP arrays is that a fixed set of loci are interrogated, and therefore novel SNPs cannot be discovered. Therefore, SNP arrays, unavoidably, result in ascertainment bias in diversity analyses when the individuals that were used for SNP selection are not representative for the diversity in the studied population.

WGS is increasingly being applied for reference genome assembly, genomic variation discovery, GWAS and genomic selection. However, there are still several challenges with WGS, including a good quality standard for genome assembly and genome annotation, as well as access to sufficient computational resources. WGS is an important and fundamental tool in aquaculture, first and foremost for SNP discovery necessary to design SNP chips. Several strategies have been used to sequence the genomes of different aquacultural species. The combination of short reads (using the Illumina sequencing platform) and long reads (using PacBio or Nanopore sequencing) provides a powerful, cost-effective way to improve the assembly quality. So far, genome assemblies for most of major aquaculture species, and high quality assemblies, resolved to chromosome level, are available, for instance, for carp (Wu et al., 2022), salmonids (Sävilammi et al., 2019), and tilapia (Tao et al., 2021).

Compared to SNP arrays and GBS, whole genome sequencing offers a huge range of additional possibilities for aquaculture species. For example, sex-determining systems are complex in fish. With GBS and SNP array technologies it is difficult to

narrow down the candidate genes and casual variants involved in sex-determination. A genome-wide association analyses based on whole-genome sequencing by Li et al. (2020) showed that 150 SNPs were significantly associated with sex, of which 76 displayed sex specificity. A major QTL, around 4 CM region mapped to LG21, conferring resistance to infectious pancreatic necrosis in Atlantic salmon, is another good example (Moen et al., 2009). Using WGS, the epithelial cadherin (*cdh1*) gene was identified as the key gene (Moen et al., 2015).

Knowing the advantages and disadvantages of each genotyping strategy, one can design a decision tree that can help decide which technology should be chosen for specific applications (shown in **Figure 2**). If a good quality reference genome is not available, the preferred option is to generate a de novo assembly (based on long read sequences). However, what matters most in the end is the research question that one wants to address and the budget that is available (**step 1**).

In **step 2**, three options (GBS, commercial SNP array and WGS) are commonly applied to identify QTLs associated with the traits of interest. If there is no good SNP array available for the target species, GBS and WGS are both good options. Since only a small portion of the genome is characterized, GBS usually will not include causal variants. Furthermore, compared to SNP arrays, GBS requires relatively complex bioinformatics analyses. The price of GBS, however, is comparable to that of SNP arrays, without, crucially, the upfront costs. WGS has been widely used to identify putative casual variants because it allows to narrow down the target region from the phenotype-genotype association. However, the main disadvantage of WGS is that samples usually need to be sequenced at relatively high genome coverages, which increases the sequencing price and requires more computational resources.

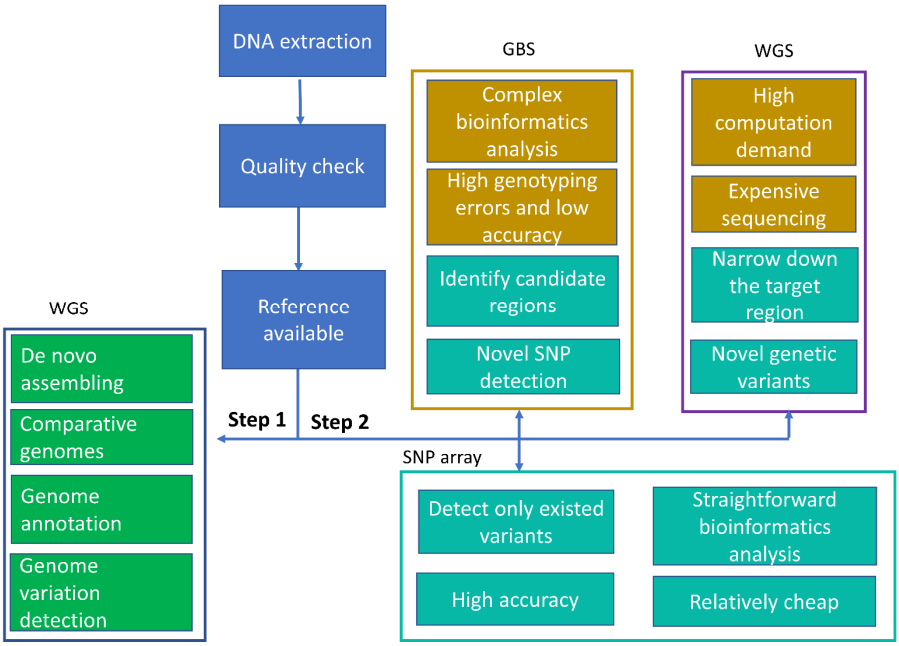


Figure 2. A framework for choosing the genomic technologies.

6.5 Concluding remarks

Challenging environments are threatening to the quantity and quality of aquaculture production. In this thesis, I characterize the genomic architecture during selection for adaptation to three crucial environmental factors: dissolved oxygen, salinity, and temperature from forward and reverse genetic approaches. I present the essential QTLs that can lead to performance divergences in a specific challenging environment. I show how QTLs interact differently across challenging environments and further lead to performance divergences. These studies highlight the essential biological processes and metabolic pathways underlying adaptation to a specific challenging environment. Furthermore, these studies provide these candidate genetic markers to selectively breed for better adapted fish. However, the current application for marker-assisted selection may be limited by the genetic variance from these markers. Therefore, I suggest the inclusion of regulatory networks (gene expression, methylation status) and environmental data in future studies.

References

Alakärppä, E., Salo, H.M., Valledor, L., Cañal, M.J., Häggman, H., Vuosku, J., 2018. Natural variation of DNA methylation and gene expression may determine local adaptations of Scots pine populations. Journal of experimental botany 69, 5293–5305.

Anttila, K., Dhillon, R.S., Boulding, E.G., Farrell, A.P., Glebe, B.D., Elliott, J.A.K., Wolters, W.R., Schulte, P.M., 2013. Variation in temperature tolerance among families of atlantic salmon (Salmo salar) is associated with hypoxia tolerance, ventricle size and myoglobin level. Journal of Experimental Biology 216, 1183–1190. https://doi.org/10.1242/jeb.080556

Bao, B., Chao, H., Wang, H., Zhao, W., Zhang, L., Raboanatahiry, N., Wang, X., Wang, B., Jia, H., Li, M., 2018. Stable, environmental specific and novel QTL identification as well as genetic dissection of fatty acid metabolism in Brassica napus. Frontiers in Plant Science 9, 1018.

Barrett, R.D.H., Hoekstra, H.E., 2011. Molecular spandrels: tests of adaptation at the genetic level. Nature Reviews Genetics 12, 767–780.

Barrio, A.M., Lamichhaney, S., Fan, G., Rafati, N., Pettersson, M., Zhang, H., Dainat, J., Ekman, D., Höppner, M., Jern, P., Martin, M., Nystedt, B., Liu, X., Chen, W., Liang, X., Shi, C., Fu, Y., Ma, K., Zhan, X., Feng, C., Gustafson, U., Rubin, C.J., Almén, M., Blass, M., Casini, M., Folkvord, A., Laikre, L., Ryman, N., Lee, S.Y., Xu, X., Andersson, L., 2016. The genetic basis for ecological adaptation of the Atlantic herring revealed by genome sequencing. eLife 5, 1–32. https://doi.org/10.7554/eLife.12081

Bergstedt, J.H., Pfalzgraff, T., Skov, P. V., 2021. Hypoxia tolerance and metabolic coping strategies in Oreochromis niloticus. Comparative Biochemistry and Physiology -Part A : Molecular and Integrative Physiology 257, 110956. https://doi.org/10.1016/j.cbpa.2021.110956

Bernard, M., Dehaullon, A., Gao, G., Paul, K., Lagarde, H., Charles, M., Prchal, M., Danon, J., Jaffrelo, L., Poncet, C., 2022. Development of a high-density 665 K SNP array for rainbow trout genome-wide genotyping. bioRxiv.

Bomblies, K., Peichel, C.L., 2022. Genetics of adaptation. Proceedings of the National Academy of Sciences 119, e2122152119.

Del Corvo, M., Lazzari, B., Capra, E., Zavarez, L., Milanesi, M., Utsunomiya, Y.T., Utsunomiya, A.T.H., Stella, A., de Paula Nogueira, G., Garcia, J.F., 2021. Methylome patterns of cattle adaptation to heat stress. Frontiers in genetics 12, 633132.

Dubin, M.J., Zhang, P., Meng, D., Remigereau, M.-S., Osborne, E.J., Casale, F.P., Drewe, P., Kahles, A., Jean, G., Vilhjálmsson, B., 2015. DNA methylation in Arabidopsis has a genetic basis and shows evidence of local adaptation. elife 4, e05255.

Elshire, R.J., Glaubitz, J.C., Sun, Q., Poland, J.A., Kawamoto, K., Buckler, E.S., Mitchell, S.E., 2011. A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. PloS one 6, e19379.

Everett, M. V., Seeb, J.E., 2014. Detection and mapping of QTL for temperature tolerance and body size in Chinook salmon (Oncorhynchus tshawytscha) using genotyping by sequencing. Evolutionary Applications 7, 480–492. https://doi.org/10.1111/eva.12147

Falconer, D.S., 1996. Introduction to quantitative genetics. Pearson Education India.

Garrison, E., Marth, G., 2012. Haplotype-based variant detection from short-read sequencing. arXiv preprint arXiv:1207.3907.

Glaubitz, J.C., Casstevens, T.M., Lu, F., Harriman, J., Elshire, R.J., Sun, Q., Buckler, E.S., 2014. TASSEL-GBS: a high capacity genotyping by sequencing analysis pipeline. PloS one 9, e90346.

Guan, L., Chi, W., Xiao, W., Chen, L., He, S., 2014. Analysis of hypoxia-inducible factor alpha

polyploidization reveals adaptation to Tibetan plateau in the evolution of schizothoracine fish. BMC evolutionary biology 14, 1–13.

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

Gutierrez, A.P., Yáñez, J.M., Davidson, W.S., 2016. Evidence of recent signatures of selection during domestication in an Atlantic salmon population. Marine genomics 26, 41–50.

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., Penman, D.J., Davassi, A., Brew, F., Tinch, A.E., Gharbi, K., Hamilton, A., 2014. Development and validation of a high density SNP genotyping array for Atlantic salmon (Salmo salar). BMC Genomics 15, 1–13. https://doi.org/10.1186/1471-2164-15-90

Jackson, T.R., Ferguson, M.M., Danzmann, R.G., Fishback, A.G., Ihssen, P.E., O'Connell, M., Crease, T.J., 1998. Identification of two QTL influencing upper temperature tolerance in three rainbow trout (Oncorhynchus mykiss) half-sib families. Heredity 80, 143–151. https://doi.org/10.1038/sj.hdy.6882890

Jamshed, M., Jia, F., Gong, J., Palanga, K.K., Shi, Y., Li, J., Shang, H., Liu, A., Chen, T., Zhang, Z., 2016. Identification of stable quantitative trait loci (QTLs) for fiber quality traits across multiple environments in Gossypium hirsutum recombinant inbred line population. BMC genomics 17, 1–13.

Jeon, G.-J., 2002. Efficiency of Marker Assisted Selection (MAS) over The Phenotypic Selection for Economic Traits in Economic Animals. Journal of Animal Science and Technology 44, 669–676.

Jin, Y., Zhou, T., Geng, X., Liu, S., Chen, A., Yao, J., Jiang, C., Tan, S., Su, B., Liu, Z., 2017. A genome-wide association study of heat stress-associated SNPs in catfish. Animal Genetics 48, 233–236. https://doi.org/10.1111/age.12482

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). Frontiers in genetics 9, 472.

Kinoshita, T., Seki, M., 2014. Epigenetic memory for stress response and adaptation in plants. Plant and Cell Physiology 55, 1859–1863.

Li, H.L., Gu, X.H., Li, B.J., Chen, C.H., Lin, H.R., Xia, J.H., 2017. Genome-Wide QTL Analysis Identified Significant Associations Between Hypoxia Tolerance and Mutations in the GPR132 and ABCG4 Genes in Nile Tilapia. Marine Biotechnology 19, 441–453. https://doi.org/10.1007/s10126-017-9762-8

Li, Y.L., Xing, T.F., Liu, J.X., 2020. Genome-wide association analyses based on whole-genome sequencing of Protosalanx hyalocranius provide insights into sex determination of Salangid fishes. Molecular Ecology Resources 20, 1038–1049. https://doi.org/10.1111/1755-0998.13172

Lillehammer, M., Goddard, M.E., Nilsen, H., Sehested, E., Olsen, H.G., Lien, S., Meuwissen, T.H.E., 2008. Quantitative trait locus-by-environment interaction for milk yield traits on Bos taurus autosome 6. Genetics 179, 1539–1546.

Lin, X., 2017. An investigation into the genetic basis of variation in hypoxia tolerance in Atlantic salmon. Master thesis 94.

López, M.E., Cádiz, M.I., Rondeau, E.B., Koop, B.F., Yáñez, J.M., 2021. Detection of selection signatures in farmed coho salmon (Oncorhynchus kisutch) using dense genome-wide information. Scientific Reports 11, 1–13. https://doi.org/10.1038/s41598-021-86154-w

Mackay, T.F.C., Stone, E.A., Ayroles, J.F., 2009. The genetics of quantitative traits: challenges and prospects. Nature Reviews Genetics 10, 565–577.

Melo, A.T.O., Bartaula, R., Hale, I., 2016. GBS-SNP-CROP: a reference-optional pipeline for SNP discovery and plant germplasm characterization using variable length, paired-end genotyping-by-sequencing data. BMC bioinformatics 17, 1–15.

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

Moen, T., Baranski, M., Sonesson, A.K., Kjøglum, S., 2009. Confirmation and fine-mapping of a major QTL for resistance to infectious pancreatic necrosis in Atlantic salmon (Salmo salar): Population-level associations between markers and trait. BMC Genomics 10, 1–14. https://doi.org/10.1186/1471-2164-10-368

Moen, T., Torgersen, J., Santi, N., Davidson, W.S., Baranski, M., Ødegård, J., Kjøglum, S., Velle, B., Kent, M., Lubieniecki, K.P., Isdal, E., Lien, S., 2015. Epithelial cadherin determines resistance to infectious pancreatic necrosis virus in Atlantic salmon. Genetics 200, 1313–1326. https://doi.org/10.1534/genetics.115.175406

Mulder, H.A., 2007. Methods to optimize livestock breeding programs with genotype by environment interaction and genetic heterogeneity of environmental variance. Wageningen University and Research.

Naval-Sanchez, M., McWilliam, S., Evans, B., Yáñez, J.M., Houston, R.D., Kijas, J.W., 2020. Changed Patterns of Genomic Variation Following Recent Domestication: Selection Sweeps in Farmed Atlantic Salmon. Frontiers in Genetics 11. https://doi.org/10.3389/fgene.2020.00264

Nayfa, M.G., Jones, D.B., Benzie, J.A.H., Jerry, D.R., Zenger, K.R., 2020. Comparing Genomic Signatures of Selection Between the Abbassa Strain and Eight Wild Populations of Nile Tilapia (Oreochromis niloticus) in Egypt. Frontiers in Genetics 11, 1132. https://doi.org/10.3389/fgene.2020.567969

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. Molecular ecology resources 15, 662–672.

Penaloza, C., Manousaki, T., Franch, R., Tsakogiannis, A., Sonesson, A.K., 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 113, 2096–2107.

Peñaloza, C., Robledo, D., Barría, A., Trinh, T.Q., Mahmuddin, M., Wiener, P., Benzie, J.A.H., Houston, R.D., 2020a. Development and validation of an open access SNP array for Nile Tilapia (oreochromis niloticus). G3: Genes, Genomes, Genetics 10, 2777–2785. https://doi.org/10.1534/g3.120.401343

Peñaloza, C., Robledo, D., Barría, A., Trịnh, T.Q., Mahmuddin, M., Wiener, P., Benzie, J.A.H., Houston, R.D., 2020b. Development and validation of an open access SNP array for Nile tilapia (Oreochromis niloticus). G3: Genes, Genomes, Genetics 10, 2777–2785.

Perea, C., De La Hoz, J.F., Cruz, D.F., Lobaton, J.D., Izquierdo, P., Quintero, J.C., Raatz, B., Duitama, J., 2016. Bioinformatic analysis of genotype by sequencing (GBS) data with NGSEP. BMC genomics 17, 539–551.

Regan, M.D., Gill, I.S., Richards, J.G., 2017. Metabolic depression and the evolution of hypoxia tolerance in threespine stickleback, Gasterosteus aculeatus. Biology Letters 13. https://doi.org/10.1098/rsbl.2017.0392

Reid, N.M., Proestou, D.A., Clark, B.W., Warren, W.C., Colbourne, J.K., Shaw, J.R., Karchner, S.I., Hahn, M.E., Nacci, D., Oleksiak, M.F., 2016. The genomic landscape of rapid repeated evolutionary adaptation to toxic pollution in wild fish. Science 354, 1305–1308.

Robledo, D., Palaiokostas, C., Bargelloni, L., Martínez, P., Houston, R., 2018. Applications of genotyping by sequencing in aquaculture breeding and genetics. Reviews in Aquaculture 10, 670–682. https://doi.org/10.1111/raq.12193

Sae-Lim, P., Gjerde, B., Nielsen, H.M., Mulder, H., Kause, A., 2016. A review of genotype-by-environment interaction and micro-environmental sensitivity in aquaculture species. Reviews in Aquaculture 8, 369–393.

Sävilammi, T., Primmer, C.R., Varadharajan, S., Guyomard, R., Guiguen, Y., Sandve, S.R., Vøllestad, L.A., Papakostas, S., Lien, S., 2019. The chromosome-level genome assembly of European grayling reveals aspects of a unique genome evolution process within salmonids. G3: Genes, Genomes, Genetics 9, 1283–1294.

Sejian, V., Bhatta, R., Gaughan, J.B., Dunshea, F.R., Lacetera, N., 2018. Adaptation of animals to heat stress. Animal 12, s431–s444.

Setyawan, P., Aththar, M.H.F., Imron, I., Gunadi, B., Haryadi, J., Bastiaansen, J.W.M., Camara, M., Komen, H., 2022. Genetic parameters and genotype by environment interaction in a unique Indonesian hybrid tilapia strain selected for production in brackish water pond culture. Aquaculture 738626.

Somorjai, I.M.L., Danzmann, R.G., Ferguson, M.M., 2003. Distribution of temperature tolerance quantitative trait loci in Arctic charr (Salvelinus alpinus) and inferred homologies in rainbow trout (Oncorhynchus mykiss). Genetics 165, 1443–1456.

Tao, W., Xu, L., Zhao, L., Zhu, Z., Wu, X., Min, Q., Wang, D., Zhou, Q., 2021. High-quality chromosome-level genomes of two tilapia species reveal their evolution of repeat sequences and sex chromosomes. Molecular Ecology Resources 21, 543–560.

Tong, C., Tian, F., Zhao, K., 2017. Genomic signature of highland adaptation in fish: a case study in Tibetan Schizothoracinae species. BMC genomics 18, 1–9.

Torkamaneh, D., Laroche, J., Bastien, M., Abed, A., Belzile, F., 2017. Fast-GBS: a new pipeline for the efficient and highly accurate calling of SNPs from genotyping-by-sequencing data. BMC bioinformatics 18, 1–7.

Vilas, R., Vandamme, S.G., Vera, M., Bouza, C., Maes, G.E., Volckaert, F.A.M., Martínez, P., 2015. A genome scan for candidate genes involved in the adaptation of turbot (Scophthalmus maximus). Marine Genomics 23, 77–86. https://doi.org/10.1016/j.margen.2015.04.011

Wakchaure, R., Ganguly, S., Praveen, P.K., Kumar, A., Sharma, S., Mahajan, T., 2015. Marker assisted selection (MAS) in animal breeding: a review. J. Drug. Metab. Toxicol 6, e127.

Wang, N., Yuan, Y., Wang, H., Yu, D., Liu, Y., Zhang, A., Gowda, M., Nair, S.K., Hao, Z., Lu, Y., San Vicente, F., Prasanna, B.M., Li, X., Zhang, X., 2020. Applications of genotyping-by-sequencing (GBS) in maize genetics and breeding. Scientific Reports 10, 1–12. https://doi.org/10.1038/s41598-020-73321-8

Wang, X., Liu, S., Jiang, C., Geng, X., Zhou, T., Li, N., Bao, L., Li, Y., Yao, J., Yang, Y., Zhong, X., Jin, Y., Dunham, R., Liu, Z., 2017. Multiple across-strain and within-strain QTLs suggest highly complex genetic architecture for hypoxia tolerance in channel catfish. Molecular Genetics and Genomics 292, 63–76. https://doi.org/10.1007/s00438-016-1256-2

Wang, X., Qu, M., Liu, Y., Schneider, R.F., Song, Y., Chen, Z., Zhang, H., Zhang, Y., Yu, H., Zhang, S., 2022. Genomic basis of evolutionary adaptation in a warm-blooded fish. The Innovation 3, 100185.

Wickland, D.P., Battu, G., Hudson, K.A., Diers, B.W., Hudson, M.E., 2017. A comparison of genotyping-by-sequencing analysis methods on low-coverage crop datasets shows advantages of a new workflow, GB-eaSy. BMC bioinformatics 18, 1–12.

Wu, C.-S., Ma, Z.-Y., Zheng, G.-D., Zou, S.-M., Zhang, X.-J., Zhang, Y.-A., 2022. Chromosome-level genome assembly of grass carp (Ctenopharyngodon idella) provides insights into its genome evolution. BMC genomics 23, 1–14.

Yu, L.-X., Zheng, P., Bhamidimarri, S., Liu, X.-P., Main, D., 2017. The impact of genotyping-by-sequencing pipelines on SNP discovery and identification of markers associated with verticillium wilt resistance in autotetraploid alfalfa (Medicago sativa L.). Frontiers in Plant Science 8, 89.

Zaïm, M., Kabbaj, H., Kehel, Z., Gorjanc, G., Filali-Maltouf, A., Belkadi, B., Nachit, M.M., Bassi, F.M., 2020. Combining QTL analysis and genomic predictions for four durum wheat populations under drought conditions. Frontiers in genetics 11, 316.

Summary

Summary

Aquaculture is a sustainable way to meet the increased human demand for animal food. However, sustainability is threatened by the increasing effect of climate change. The most important threats to aquaculture resulting from climate change are high water temperatures that fall outside of the physiological tolerance range, increase of and rapid change in salinity due to drought and flooding, acidification of seawater due to carbon dioxide uptake, and last but not least oxygen shortage caused by any combination of the above that could result in algal blooms. Although a reasonable genetic gain can be achieved by traditional selective breeding, aquaculture may not grow sufficently due to the complexity of traits underlying adaptation to these environmental stressors. However, the development of genomic technologies enabled us to investigate the genomic architecture of traits of interest. The aim of this thesis is to provide an insight into the genomic architecture of adaptation to challenging environments of aquaculture species under farming conditions and the underlying essential biological processes and metabolic pathways.

In **Chapter 2** we used two groups of the Genetic Improvement of Farmed Tilapia strain that originated from the same genetic background. The Nile tilapia were raised in aerated (normoxic) and non-aerated (hypoxic) ponds, respectively. We observed different genomic architectures associated with early and later growth in hypoxic and normoxic environments, indicating a transition in metabolism. We highlighted that MAPK and VEGF signalling are significantly involved in the regulation of later-stage growth under hypoxia. Furthermore, by a meta-analysis of GWAS results, we identified thirty-three SNPs that are significantly associated with growth across two environments, while the genes linked to these SNPs were mostly involved in nervous system development, organ growth and oocyte maturation. These findings suggest a possible shared effect independent of the two environments.

In **Chapter 3**, critical swimming performance (U_{crit}) is studied as a potential indicator trait for hypoxia tolerance. Fish with higher U_{crit} are not only supposed to perform better in cardio-respiratory health, but also to cope with hypoxia better. We found slightly negative correlations between swimming performance in early life and growth traits based on the genomic relationship matrix, suggesting that fish with better swimming performance have a slower growth tendency later in life. We also identified the genomic architecture associated with U_{crit} early in life and identified several potential target genes such as *hip1*, *hectd1*, *elna*, *smyd1b*, *rrp12* and *pprc*. Moreover, a clear pleiotropic effect of three SNPs was found between swimming performance and growth traits, while the remaining six SNPs only affect swimming

performance, but not growth. The genetic markers identified in the study can be used to select fish with cardio-respiratory health and growth that can potentially grow well in a hypoxic environment.

The results described in **Chapter 4** are based on a saline-tolerant and highly productive tilapia strain that was developed by the aquaculture research institute and named "Sukamandi". We show that the Sukamandi genome is predominantly of Nile tilapia origin, with about 9% of the genome derived from blue tilapia. We further show that eight salinity tolerance genes, including *caprin1a*, *nucb2a*, *abcb10*, *slc12a10.1*, *cacna1ab*, *ulk2*, *slc25a24* and *cdh1*, were strongly selected after a four-generation breeding program. Moreover, the genomic architecture that was introgressed from blue tilapia also confers the observed salinity tolerance. Collectively, the results help us understand fish adaptation in a saline envionment.

In **Chapter 5** we study the same genetic background of gilthead seabream that grew out in two distinct commercial sites (Spain and Greece). We describe a region between 19.66 Mb to 46.84 Mb on chromosome 22 that is strongly associated with growth traits in Spain. However, none of the SNPs within this region are associated with growth in Greece. Moreover, a similar pattern was found for organ weight, suggesting a divergent genomic architecture under genotype-by-environment interaction. By performing gene ontology and KEGG enrichment analyses for the target QTLs region, we show that cell adhesion, light absorption, and ECM-receptor interaction are the most prominent terms related to growth in Spain but not in Greece. Also, we show that ATP binding activity and response to stimulus are more prominent processes for organ weight in Spain than in Greece. The QTLs identified in this study provide insight into the genomic basis of fish growth traits and organ weight under a temperature challenging environment.

In **Chapter 6**, the general discussion, I discuss the genomic basis of environmental adaptation using forward and reverse approaches, suggesting a schematic diagram of combining two approaches to study the genetic mechanism of adaptation. To further strengthen our understanding of genotype-by-environment interaction, I discuss QTL by environment interaction from a genomic perspective. I also discuss the advantages and disadvantages of current genomic technologies from SNP array, genotype-by-sequencing to whole genome sequencing used in the studies described in this thesis. Lastly, I give concluding remarks and suggest the inclusion of regulatory network (gene expression, methylation status) and environmental data in future studies.

摘要

水产养殖是人类对日益增长食物需求的一种可持续生产方式。然而，受全球气候变化的影响越来越大，可持续发展受到威胁。全球气候变化对水产养殖的威胁主要表现在：超出生理耐受范围的变化，包括全球变暖相关的水温，干旱和洪水导致的水盐度，二氧化碳导致的水酸化，以及这些组合与水藻共同导致的水溶解氧的方面。虽然传统的选择育种能够取得一定的遗传进展，但因适应这些环境压力具有潜在的复杂性，水产养殖在短期内无法实现可持续增长。随着基因组技术的发展，为我们解析重要经济性状的遗传机制提供了可能性。本论文旨在揭示水产养殖物种在养殖条件下适应具有挑战性环境的基因组结构，以及潜在重要的生物过程和代谢途径。

在第二章节中，我们在不同水溶解氧的环境下，养殖了两组具有相同遗传背景的尼罗罗非鱼。我们观察到不同的基因分别调控低氧和常氧环境下的罗非鱼前期生长和后期生长，推测这可能是新陈代谢的转变。其中，MAPK 和 VEGF 代谢通路显着参与了低氧条件下尼罗罗非鱼后期生长的调节。此外，通过对 GWAS 结果的荟萃分析，我发现 33 个 SNP 位点与低氧和常氧环境下罗非鱼生长都显著相关，相关的基因主要涉及神经系统发育、器官发育和卵母细胞成熟，这可能表明它们是独立于两种环境的共享效应。

在第三章节中，我们研究了鱼低氧耐受性的指标 – 游泳能力。鱼的低氧耐受性越强，游泳能力就越强，并且心肺功能更好。基于遗传相关性分析，我们发现罗非鱼早期的游泳性能和生长速度之间存在负相关，这表明游泳能力较好的罗非鱼，呈现出较慢的生长趋势。我们进一步解析了与罗非鱼游泳能力相关的重要遗传信息，其中关键的调控基因有 *hip1*、*hectd1*、*elna*、*smyd1b*、*rrp12* 和 *pprc*。此外，我们还发现 3 个 SNP 位点与游泳性能和生长性状存在显著的多效性，而另外 6 个 SNP 位点只影响游泳性能，不影响生长性状。本研究中确定的遗传标记可用于选择心肺健康和生长状况良好的鱼类，这些鱼类可能在低氧环境中生长良好。

在第四章节中，印尼水产研究所培育出了一种耐盐性好且高产的罗非鱼，名为“Sukamandi”。通过比较基因组分析，我们发现 Sukamandi 基因组主要与尼罗罗非鱼同源，而只有 9%的基因组来自蓝罗非鱼。经过四代选育，我们发现与 Sukamandi 耐盐性相关的 8 个基因受到很强的选择，包括 *caprin1a*、*nucb2a*、*abcb10*、*slc12a10.1*、*cacna1ab*、*ulk2*、*slc25a24* 和 *cdh1*。此外，蓝罗非鱼渗入的基因组部分也贡献了 Sukamandi 的耐盐性。这些结果有助于我们理解鱼类对高盐环境的遗传适应机制。

在第五章节中，我们在不同水温的渔场（西班牙和希腊），养殖具有相同遗传背景的金头鲷。通过基因组定位，我们发现 22 号染色体上 19.66 Mb 到 46.84 Mb 遗传区间与西班牙渔场金头鲷的生长性状显著相关。但这个遗传区间对希腊渔场金头鲷的生长并无影响。此外，我们发现金头鲷在器官发育上也有同样的规律。这表明基因在与环境的互作下共同决定表型。对调控表型的相关基因进行功能富集，我们发现细胞粘附、光吸收和 ECM 受体互作等通路对西班牙渔场金头鲷的生长起到显著的调控作用，但对希腊渔场则不然。而 ATP 结合活性和应激反应在西班牙渔场比希腊渔场对器官发育调控更重要。本研究从基因组的角度解析了，在不同水温度刺激环境下鱼类生长和器官发育的遗传机制。

在第六章讨论章节中，本人系统的比较了正向与反向的遗传研究方法，在解析鱼类环境适应性遗传机制的应用，并提出如何进一步将两种方法结合起来，从而更好的解析鱼类环境适应性机制。此外，为了进一步理解遗传与环境的互作，本人从基因组的角度探讨了基因与环境互作关系。我还探讨了当前与性状解析相关的主流基因组技术 （包括基因芯片，简化基因组测序以及全基因组测序）的优缺点。最后，本人给出了全文结论以及下一步的研究方向，以更好的应对因全球环境变化而带来的水产养殖挑战。

Appendices

Acknowledgements

As time goes by, the life of being a PhD student draws to a close. Four years seems like a long time before starting, but it feels very short after really adapting to the life here. Although the journal contains challenges, suffering and hardworking, it also contained progress, achievement and happiness. Certainly, I could not go through all and reach this milestone without receiving support from many people. In this section, I would like to express my gratitude to those people who helped me along the way.

First of all, I would like to sincerely thank my supervision teams: **Martien**, **Hans** and **Hendrik-Jan**. During the whole PhD journey, you trained me to be an independent scientist who has a view with problems and analyze skills. You all enough me when encountered all the difficulties during research. Dear **Martien,** I will not start this PhD study without you. I still remembered the first time I contact you through email when I was fulfilled with uncertainty and worries about the future. You cannot imagine how happy and excited I was when received the invitation from you. I was impressed by your knowledge the first time met you in Beijing during a scientific meeting. I also enjoy each discussion we had during our meetings, and you always try to train me to be critical and not to over-interpret the results I found with analysis. I appreciate having you on the team and giving me feedback whenever I need it! Dear **Hans**, I know you after my application for CSC scholarship was granted. I have to say you play an incredibly important role in my PhD more than I had thought. I like the way how you communicate with me – discussing difficulties and struggling problems in a relaxing and pleasant way. As an expert in fish breeding and genetics, you keep challenging me to think about what the results explained underneath and how they connect the aquaculture practical. I was impressed by your big view and storytelling skill, especially during our intensive meetings this Summer. Dear **Hendrik-Jan**, you have generously helped me from the topics selection, data analyses, scientific writing to thesis finalizing. Every conversation with you has benefit me a lot during the last four years. I was impressed with your broad knowledge of both science and society. After our 'formal' meeting, we are more like a friend to share our opinions about recent facts and news. I was impressed by your suggestions on how to balance between career and family. Without you all three, I cannot finalize my thesis as far as I had. Thanks again, my dear supervisors.

To **Han** and **John (Bastiaansen)**, you are the special expert next to my supervision team in this thesis. **Han**, I was impressed by your in-depth understanding of Genotype-by-environment and quantitative genetics models. **John**, you are not one

of my supervision team. But you are always available to discuss my results and difficulties. You helped me a lot with methodologies formalization. I am especially impressed with your thorough consideration with things. Thanks also to **John (Benzie)** and **Trong** for making several experiments becoming possible in WorldFish, and John's encouragement to train me as an independent journal reviewer.

I also thank people from the aquaculture discussion group. **Mark**, you join our 'tilapia' group during the middle of my PhD. You are the person who attended the most for the tilapia student meeting each week, to make sure everything is going in the correct direction with your 'nephew'. Although I had a bit tough time at the beginning, you taught me the importance of teamwork. I also enjoy the discussion with you about scientific writing and the tips you gave. **Arjan**, it was a nice experience to collaborate with you and receive your fruitful feedback on the chapter about swimming performance. **Wout**, thank you for helping me design the indoor salinity experiment. Although it did not include in this thesis due to the delay caused by Covid, I am sure it will a nice scientific article. **Farid** and **Priadi**, we are the VIP members of the tilapia student meeting. I enjoy your discussion and your help with sampling organization and experiment design. I also enjoy the drink and talks we had on many occasions. **Samuel** and **Fasil**, who come from the same country (Ethiopia) and defense their PhD thesis on the same day. **Samuel**, my best brother and officemate from Africa, I received endless help from you whenever I ask. We studied and did sports together at the beginning year of my PhD. And we still manage to share our progress and life whenever you are. A big thanks for your hard work and contributions to the experiments in this thesis. **Fasil**, thank you for being a patient teammate during the Genomics course and your optimism always impressed me. Thanks to the other fish guys, **Benan**, **Wisdom**, **Yukko**, **Pauline and Meilisa** (the fresh PhD candidate)**,** who constitute a pleasant atmosphere at work.

My thanks go further to colleagues from the Genomics interest group every Monday. **Ole**, **Richard**, **Mirte**, **Martijn** and **Marta**, you guys are golden members there to be ready to spread your knowledge and give useful advice whenever I presented in the Genomics meeting. A special thanks to **Jani** and **Carolina** for being my paranymphs. **Carolina**, I am happy to be your friend and do all the funny activities (cycling, kickboxing, natural sporty, beach volleyballs...) together with you, which really enrich my spare life in the Netherlands. **Jani**, I like your open mind and our talk during our coffee time. I can see how fast your knowledge gain during the last three years. I am sure you will continue to gain in the near future. **Annemieke** and **Gibbs**, we are supervised by the same daily supervisors, I enjoy a lot gaining my python

programming knowledge with you guys (including **Jani** here) during the Covid isolated time. **Rayner**, thanks for your help during my analysis and the happy time when we played badminton, tennis and beach volleyball, special thanks to teach me movement of Latin dance. Thanks also to other Genomics members including **Magi Pedro**, **Marije**, **Jeroen** for your companionship. There are colleagues that I cannot put in a particular category above but still play an important role in my PhD. **Harmen, Henri**, **Lim**, **Ibrhim**, **Renzo**, **Malou**, **Pascal**, **Marieke**, **Biaty**, **Chiara**, who are graduated PhD and motivated me to progress now and then. I would also like to thank **Fatma**, **Lucas**, **Erik**, **Anouk**, **Matias** and **Gerbich** for the party and all the fun moments that color my life. I also thank to **Lisette**, **Paddy** and **ABG secretary** for their great help and assistance during my PhD.

Thanks to the Chinese friends who derserved to be mentioned here. **Langqing (琅青)** and **Zhou (吴舟)**, thanks to all discussion from genomics, bioinformatics to life. **Shuwen (淑雯)**, **Qiuyu (秋雨)**, **Haibo (海博)** and **Yebo (业博)** and thank to your help to solve my questions related to all kinds of questions in life. **Yun (喻运)** and **Xue (潘学)**, we started our PhD almost the same time, and defense our PhD with only a few months difference. I am sure that we have learned a lot from each other and motivated our progress. **Zhuoshi (卓识)**, **Wenye (文晔)**, **Aixin (爱心)**, **Ziwei (紫薇)**, **Rui (师瑞)**, **Qitiong (其通)**, **Liyan (立言), Zhen (潘珍)**, **Yubo (钰渤)** for filling in my blank moments with pleasures and happiness, especially during traditional Chinese festivals.

Junnan (俊楠), my sister from Henan. The first time we knew each other was through a call in the last year of my bachelor in Nanchang. I received fruitful advice on how to prepare for the entrance exam of Master. You made me the first attempt at something that looks impossible, and the same for PhD application later. I am happy to see your successful PhD defense a few month ago. So proud of you! **Chuzhe (春哲)**, we travelled to the Netherlands from China at the same day. I am happy to see the growth of our strong friendship during the last four years. We have witnessed our development both in research capabilities and thought. I wish you a sweet life together with **Xiaochen** (**晓晨**). **Liuchen (刘臣)**, **wangqi (王琦)** and **jinxiao (金晓)** thank for all the Chinese party and the well-organized travel we had, which really means a lot to me. Special thanks to **Lei (邓磊)**, who help me to practice my oral speaking during preparing IELTS. Thank **Fang (吕芳)**, **Yuhuang (玉煌)**, **Yue (王越), Dengke (登科)**, **Mengting (梦婷)**, **Yixin (义信)**, **Yaqing (雅清)**, **Shujuan (淑娟)** for sharing your delicious food and enjoyable time we had.

I am a big fan of sports, so I would like to thank a few people. **Lei (张磊)**, **Yuqi (喻琦)**, **Jiayu (嘉宇)**, **Wenjie (文杰)**, **Yan (张岩)**, **Mike** from the Basetball teammates, and **Zulin (祖麟)**, **Chengcheng (成成)**, **Yajing (雅晶)**, **Xiaolin (肖林)**, **Jianing (佳宁)** from the Badminton. As the old saying, life is a movement, I hope we can keep exercising no matter how old we are.

I would like to acknowledge the China Scholarship Council for providing me with financial support. And my Master supervisors **Prof. Li (李奎教授)** and **Prof. Tang (唐中林研究员)** fully support me to study abroad. And also my sincere fellows from the same group, including **Lijing (白立景师姐)**, **Yalan (杨亚岚师兄)**, **Guanglin (钮广林师兄)** and Zishuai **(王子帅师兄)**.

Furthermore, a special thanks to **Loes**, **Egbert** and **Naghme**, you are like my family members from Wageningen, I enjoy the time with you very much and all the journeys during exploring the Netherlands.

Looking forward seeing you all in the future, please forgive me if I forgot to mention you above!

最后，我还要特别感谢我的父母，在我求学生涯中，给予我无尽的鼓励与支持。正是有了你们良好的教育，我才能走到今天。还有我亲爱的姐姐和姐夫们，有了你们无微不至的照顾，才让我能顺利完成了博士学业。特别感谢我的女朋友张秀玲，你的陪伴让我的生活变得很充实，期待与你的相聚！

About the author

Xiaofei Yu was born on 18th December 1992 in Jiangxi, China. He showed strong interest in animal when he was a kid. After graduated from Jiujiang No.1 Middle School, Xiaofei studied animal science in Jiang Xi Agriculture University and obtained his bachelor degree in 2015, for which he also did a internship at Twins Group company. Later, he succeed to continue his mater study in Chinese Academy of Agricultural Sciences under the supervision of Prof. Kui Li and Prof. Zhonglin Tang. In his MSc research project, he first time knew the high-throughput sequencing data and performed analyses under Linux command line. In July of 2018, he finished his MSc thesis entitled "Identifying LncRNAs Associated with Skeletal Muscle Growth and Development Based on Integrative Analysis of RNA-seq Data in Pigs". In October 2018, he started as a PhD candidate at Animal Breeding and Genomics group, Wageningen University & Research. The results of his PhD research are presented in this thesis entitled "Genomic architecture of selection for adaptation to challenging environments in Aquaculture" under supervision of Prof. Hans, Prof. Martien and Dr. Hendrik-Jan Megens.

(**Contact for more information**: xiaofei_yu1992@163.com)

List of publications

2022 **Yu X.,** Mengistu S.B., Mulder H.A., et al. Quantitative trait loci controlling swimming performance and their effect on growth in Nile tilapia (Oreochromis niloticus), ***Aquaculture*** 738522.

2022 **Yu X**., Setyawan P., Bastiaansen W.M. J., et al. Genomic analysis of a Nile tilapia strain selected for salinity tolerance shows signatures of selection and hybridization with blue tilapia (*Oreochromis aureus*), ***Aquaculture*** 738527.

2021 **Yu X**., Megens HJ., Mengistu S.B., et al. Genome-wide association analysis of adaptation to oxygen stress in Nile tilapia (*Oreochromis niloticus*), ***BMC Genomics*** 22, 426.

2018 **Yu X**., Wang Z., Sun H., Niu G., Li K & Tang Z. Identify Long non-coding MEG3 as candidate for skeletal muscle development and meat production traits in pigs, ***Animal Genetics*** 49(6):571-578.

2015 **Yu X**., Liang R., Li X., Sun Y., Yan D., Sheng Z., Zhou R. Characterization and analysis of blood physiological and biochemical indexes in Mountain Black Pig (in Chinese). ***China Swine Industry*** Z2,44-47.

2014 Sheng R., **Yu X**., Chen X., Wu T, CHEN J., Huang W. Exploration on the technology of establishing a multi-king complex by the new queen of Italian honeybee (in Chinese). ***Journal of Bee*** *10*,17-18.

2022 **Yu X.**, Gulzari B., Bastiaansen W.M. J., et al Genome-wide association analyses reveal genotype-by-environment interactions of production and organ weights in gilthead seabream (*Sparus aurata*). ***To be submitted***.

2022 **Yu X.**, Setyawan P., Bastiaansen W.M. J., et al. Short- and long-term salinity acclimation effect on gene expression and their genetic basis in a saline-tolerant tilapia. ***In preparation***.

Training and Supervision Plan (TSP)	**Graduate School WIAS**

A. The Basic Package	year
WIAS Introduction Day **(mandatory)**	2018
Introduction Course On Essential Skills (Frank Little)	2019
WGS course Ethics and Animal Sciences	2020
WGS course Scientific Integrity	2020

B. Disciplinary Competences	year
MSc course on Genomics	2018
Getting started in ASReml	2019
De Novo Assembly workshop	2020
PhD discussion group (ABG weekly genomics)	2018-2022
Writing research Proposal	2019
Analysis of bulk RNA-seq data	2021

C. Professional Competences	year
The essentials of scientific writing and presenting	2019
Research Data Management	2019
Project and Time Management	2018
Brain Training	2019
WAPS Council representative on WIAS Education Committee	2019-2020
Start to Teach	2019
Presenting with Impact	2020
Scientific Artwork, Data visualization and Infographics with Adobe Illustrator	2020
Paper review (two papers)	2022
The Final Touch: Writing the General Introduction and Discussion	2022

D. Societal Relevance *(recommended)*	year
Communication with the Media and the General Public	2021

E. Presentation Skills *(maximum 4 credits)*	year
WIAS Science day (Oral)	2021
WIAS Science day (Poster)	2022
European Aquaculture 2021 (Oral)	2021
WCGALP 2022 (Oral)	2022

F. Teaching competences *(max 6 credits)*	year
Supervising Genomics (ABG-3036)	2019-2021
Supervising MSc major thesis	2020

Education and Training Total (minimum 30 credits)*	**40.5**

The research described in this thesis was financially supported by Wageningen University & Research, the Koepon Foundation and WorldFish.

Xiaofei Yu was sponsored by the China Scholarship Council.

Cover design by Xiaofei Yu and Houmin Liu

Printed by: DigiForce | Proefschriftmaken.nl