博士論文

Genetic basis and genetic improvement of heterobothriosis resistance and growth performance in the tiger pufferfish, *Takifugu rubripes*, fed with low fishmeal diet

（低魚粉飼料給餌下におけるトラフグのヘテロボツリウム症耐性と成長能力に関する遺伝基盤と選抜育種に関する研究）

リン　コケツ

林 子杰

Lin Zijie

Contents

Introduction

Fishmeal shortage for aquafeeds and its possible solutions

Fishmeal produced from wild-captured fish is an important component of diets used in the aquaculture industry. However, the rising price of fishmeal and the necessity for sustainable fishing have encouraged the aquaculture industry to explore alternative protein resources (Hua et al., 2019). To reduce the amount of fishmeal, a wide range of potential replacements has been investigated, such as plant meal, animal byproducts, fishery and aquaculture byproducts, and insect meal. These have been tested on a variety of aquaculture species, e.g., Atlantic salmon (*Salmo salar*) (Belghit et al., 2019; Caballero-Solares et al., 2018a; Davidson et al., 2016; Egerton et al., 2020), Japanese flounder (*Paralichthys olivaceus*) (Kikuchi et al., 1993), rainbow trout (*Oncorhynchus mykiss*) (Adelizi et al., 1998; Rimoldi et al., 2021; Satoh et al., 2003; Yoshitomi et al., 2007), yellowtail (*Seriola quinqueradiata*) (Ido et al., 2021; Murashita et al., 2019), red sea bream (*Pagrus major*) (Seong et al., 2019), European sea bass (*Dicentrarchus labrax*) (Kaushik et al., 2004; Magalhães et al., 2017; Rimoldi et al., 2020; Serradell et al., 2020), and white shrimp (*Litopenaeus vannamei*) (Hernández et al., 2008; Motte et al., 2019; Tan et al., 2005; Xie et al., 2016, 2014).

To further reduce the environmental footprint of the aquaculture industry, single cell protein biomass from microorganisms such as bacteria, yeasts, and microalgae have been suggested as sustainable and cost effective replacements for fishmeal (Cottrell et al., 2020; Gamboa-Delgado and Márquez-Reyes, 2018; Matassa et al., 2016). Bacteria and yeasts have high protein contents and excellent nutritional profiles, which can be additionally improved by adjustment of culture methods or genetic engineering (Agboola et al., 2021; Wang et al., 2020). Recently, the use of bacteria and yeast-based diets as major protein ingredients has been explored in Atlantic salmon (Couture et al., 2019), rainbow trout (Roques et al., 2018), red sea bream (Takii et al., 2004), and white shrimp (Chen et al., 2021; Zhao et al., 2017). The availability of low fishmeal (LFM) diet would be largely depending on species specific characteristics of digestive system (Hua et al., 2019). In the worst case, LFM diets may diminish production traits (e.g., growth performance and disease resistance) of farmed fish due to the gene-environment interaction, that is, genes may have a different response under a different environment (Zhang and Belsky, 2020). On the other hand, it is highly expected that the LFM tolerance (i.e., production traits under the dietary treatment of LFM diet) can be genetically improved by means of selective breeding, as a complementary strategy for the LFM diet formulation to solve the problem of fishmeal shortage (Hua et al., 2019). Some preliminary studies documented success of genetic improvement for growth-related traits under the plant-based LFM diet, such as rainbow trout (Callet et al., 2017; Miura et al., 2020) and Amago salmon (*Oncorhynchus masou ishikawae*) (Yamamoto et al., 2015, 2016). However, the feasibility of selective breeding for the LFM tolerance has not been systematically examined.

Selective breeding in aquaculture

Selective breeding is the process to selectively improve particular traits through recurrently mating high-potential individuals and producing genetically superior progenies. This cumulative genetic gains bring about high economic returns (Oldenbroek and van der Waaij, 2015). So far, selective breeding significantly

accelerates the food production of farm animals and major crops, while its application and progress in aquaculture lags far behind, i.e., only 10% of aquaculture production is derived from selective breeding programs in 2012 (Gjedrem et al., 2012). Fortunately, for aquaculture species, the genetic improvement of economically important traits usually have higher genetic gain due to the high fecundity and genetic diversity compared to the terrestrial animals (Gjedrem and Baranski, 2009). Moreover, the production traits (i.e., growth performance, disease resistance, etc.) of aquaculture species commonly show moderate or high heritability, suggesting high potential for genetic improvement (Elaswad and Dunham, 2018; Gjedrem et al., 2012; Gjedrem and Rye, 2018; Hosoya et al., 2017). Consequently, the benefits of selective breeding have been widely recognized and it is routinely practiced in several species, such as, Atlantic salmon, rainbow trout, and Nile tilapia (*Oreochromis niloticus*) (reviewed in Houston et al., 2020). However, systematic selective breeding programs is still under development or even not existing for most farmed fish diminishing the farming efficiency (Gjedrem and Baranski, 2009). Considering the rapidly growth of aquaculture industry, cost-efficient selective breeding programs are highly demanded to establish the elite fish breeds for the most of fish species (Gjedrem et al., 2012).

Selection methods

To initiate a cost effective selective breeding program, it is essential to choose an optimal selection method, including mass selection, pedigree-based selection, marker-assisted selection and genomic selection (Oldenbroek and van der Waaij, 2015). While mass selection (also known as phenotypic selection) has been practiced since early prehistory, the first scientific attempt of mass selection was done by Robert Bakewell for terrestrial animals in the 18th century (Frana, 2003). Mass selection is based solely on phenotypes of the target traits (e.g., body size). Phenotypic values are not only the measurements of traits but also the consequences of gene-environment interaction (Oldenbroek and van der Waaij, 2015). Thus, phenotype can be further considered as an approximation of the genetic performance of the progeny. Although mass selection outperforms in cost efficiency, this method is ineffective for traits with low heritability because the genetic variance and environmental variance of the targeted traits are not differentiated. And worse still, mass selection lacks in detailed information for inbreeding control (Bentsen and Olesen, 2002).

With the developments of population genetics in the early 20th century, the pedigree-based selection was pioneered by Sewall Wright and Jay Laurence Lush (Gjedrem and Baranski, 2009), and further sophisticated by Charles Roy Henderson who developed the linear mixed model equations to solve best linear unbiased predictions (BLUP) of breeding values (Henderson, 1976, 1953). Pedigree-based selection trains a linear mixed model using phenotypes and pedigree information to estimate breeding values of each individual, enabling breeders to rank the candidate animals for selection. Compared to mass selection, pedigree-based selection is advantageous in inbreeding control and selection accuracy as this method can explicitly separate the phenotype into environmental component and genetic component (breeding value). Pedigree-based breeding methods have contributed to aquaculture development by improving economically important traits, as seen in the salmonids and tilapias (K. Janssen et al., 2017; Neira, 2010; Rye et al., 2010). However, pedigree-based methods have innate drawbacks where it is assumed that estimated breeding values of target traits for candidate individuals are the average breeding values of parents, ignoring stochastic Mendelian segregation (Mendelian sampling) within families (Gjedrem and Baranski, 2009). Thus, pedigree-based methods can not differentiate

estimated breeding values among full sibs. Prediction using large-scale pedigrees including many full- and half-sibs can solve this problem (Walsh and Lynch, 2000). However, collecting large-scale pedigree information is time-consuming and laborious especially in the practice of aquaculture because larvae are too small for tagging and thus it is necessary to keep each family in separate until fish reaches a body size large enough for tagging.

With the advent and development of DNA-based genetic markers, the association between traits of interest and those genetic material are detectable, and thus marker-assisted selection (MAS) become feasible (Wakchaure and Ganguly, 2015). This method is effective when the target trait is determined by a few loci (e. g. quantitative trait loci (QTL)) with large effects (Lande and Thompson, 1990); at least two strains have been established by means of MAS, i.e., lymphocystis resistant Japanese flounder (*Paralichthys olivaceus*) (Fuji et al., 2007) and infectious pancreatic necrosis resistant Atlantic salmon (Moen et al., 2015). However, it is often hard to find DNA markers that can explain a high proportion of genetic variance, since most of economic traits are polygenic and have complex genetic architecture in aquaculture species (Goddard and Hayes, 2009).

In 2001, Meuwissen *et al* proposed genomic selection (GS) to estimate the genomic estimated breeding values (GEBVs) of selection candidates by harnessing whole-genome high-density markers and advanced regression methods (Meuwissen et al., 2001). In GS, those markers are effective in handling errors due to Mendelian sampling by capturing genetic variance at DNA levels. Thanks to the recent advances in the next-generation sequencing (NGS) technologies, it is now affordable to genotype genome-wide single nucleotide polymorphisms (SNPs) for GS in aquaculture breeding programs (Robledo et al., 2018b). As expected, the greater performance of GS over the pedigree-based method in prediction and inbreeding control has been demonstrated by empirical studies (Tsai et al., 2015; Vallejo et al., 2017). Recently, the potential of GS for disease resistance has been seen in amoebic gill disease in Atlantic salmon (*Salmo salar*) (Robledo et al., 2018a), bacterial cold water disease resistance in rainbow trout (*Oncorhynchus mykiss*) (Vallejo et al., 2017), viral nervous necrosis in European sea bass (*Dicentrarchus labrax*) (Palaiokostas et al., 2018), and photobacteriosis in gilthead seabream (*Sparus aurata*) (Palaiokostas et al., 2016). Although the majority of the current breeding programs in aquaculture is still in its infancy, GS is highly expected to accelerate the establishment of high-performance strains (Hosoya et al., 2017).

Demand for genetic improvement in the farmed tiger pufferfish

The tiger pufferfish, *Takifugu rubripes*, is a delicacy and one of the most valuable marine fish species in Japanese aquaculture, ranking fourth in production value among cultured finfish (Hosoya et al., 2014; Ogawa, 2016). The farming efficiency of this species was largely improved by the technology for artificial fertilization developed since 1990s (Miyaki et al., 1998). However, systematic selective breeding program has not been practiced and high-performance fish breeds are not available until recently (at the start of my Ph.D. program). As the current standard feed of the tiger pufferfish containing high level of animal protein (70~80%) mainly from the fishmeal, genetic improvement of production traits under a LFM diet is highly expected to further improve farming efficiency and sustainability of the tiger pufferfish farming (Kikuchi, 2006).

Apart from growth performance, the disease resistance is an important production trait, as disease outbreaks easily hamper the aquaculture industry. One of the most serious diseases in the tiger pufferfish farming is heterobothriosis, which is a gill disease caused by a monogenean parasite *Heterobothrium okamotoi*.

The most severe infection occurs at early phases of production, just after transfer from land-based hatcheries to sea cages (Ogawa, 2002; Ogawa and Inouye, 1997). These naïve juveniles are afflicted by the parasite, persistently present at oceanic aquaculture sites, resulting in retarded growth and high mortality rate (Shirakashi et al., 2010). Therefore, frequent drug treatments are needed to control the parasitosis, which leads to high financial costs, environmental contamination and emergence of drug-resistant parasites (Ogawa, 2016). Instead of the drug treatments, genetic improvement of heterobothriosis resistance in farmed fish is considered as a cost-efficient solution. To establish selective breeding program for the resistance trait, it is essential to know the genetic basis of the trait to choose selection methods while mechanism of host immune response to *H. okamotoi* are still unclear (Igarashi et al., 2017; Matsui et al., 2020). Although the QTL analysis using a full-sib family of F_2 hybrids between a wild female grass pufferfish, *T. niphobles* and a wild male tiger pufferfish suggested that host resistance to heterobothriosis is polygenic (Hosoya et al., 2013), the genetic basis of heterobothriosis resistance in the tiger pufferfish is still unclear. Moreover, the diet change may reshape the genetic basis of these production traits because genes related to the production traits may have different response under different dietary treatments, i.e., genotype–environment interaction. As most of the disease resistance traits have moderate or high heritability in fish species (Ødegård et al., 2011; Robledo et al., 2018a) as well as the growth-related traits (Houston et al., 2020), it is expected that GS can also be applied with these traits in the species. In addition, the genetic resources of this species give a huge advantage to apply GS breeding program: its compact genome (around 400 Mbp) allows GS breeding program with fewer SNPs to reach the same density across the whole-genome compared to the other species; a high-quality genome assembly (FUGU5) simplify SNP panel construction and NGS analysis (Kai et al., 2011; Sato et al., 2019).

Objectives

In this study, my objective is to investigate the genetic basis of disease resistance and body size and the feasibility of selective breeding by means of GS under the dietary treatment of LFM diet using the tiger pufferfish.

In Chapter 1, I first investigated the genetic basis of the two traits and possibility of simultaneous improvements using a small-scale experiment. Then, in Chapter 2, I have enlarged the population size and increased SNP density to examine the generality of previous results and optimal SNP density for accurate GS for these traits. In Chapter 3, I have evaluated the effects of four types of LFM diets on growth performance, blood chemistry, transcriptomic responses in the liver, and resistance to heterobothriosis and determined the best LFM diet. In Chapter 4, I have investigated the genetic basis and feasibility of GS for these traits under the short-tern dietary treatment of the best diet. And finally, in Chapter 5, I have studied the impact of long-tern dietary treatment of LFM diet on genetic basis and feasibility of GS for body size.

Chapter 1. Availability of genomic selection for heterobothriosis resistance and body size under a standard feed

In this chapter, I have investigated heritability, genetic architecture, and the availability of genomic selection (GS) for heterobothriosis resistance and body size in the tiger pufferfish reared with a standard diet using a small population. Heritability (narrow sense) is the ratio of the additive genetic variance to the phenotypic variance, and if the trait is largely determined by the genetic factors, it means the trait have high heritability (de los Campos et al., 2015). Therefore, heritability has been widely recognized as an indicative of the potential for genetic improvements of the target trait (Mathew et al., 2018; Visscher et al., 2008). High heritability, however, does not indicate the existence of large effect genes underlying inter-individual phenotypic differences. Moreover, estimation of heritability does not refer to the underlying genetic architecture, i.e., the number of the genes, the location of the genes on the genome, and the effects of the genes affecting the trait.

The genetic architecture of the target trait is an important factor for the choice of selection method. As the extent of heritability and the number of genes affecting the trait are uncorrelated, genetic architecture should be studied separately from heritability. The first study which reveals genetic architecture of a trait is Mendel's work on pea genetics. The traits which Mendel treated (e.g., pea shape and colors) were binary traits and controlled by single (or a few) large-effect gene(s). It is relatively easy to determine the genetic architecture or detect the genomic position of the causative loci (quantitative/qualitative trait loci, QTLs) of such monogenic traits (also known as Mendelian traits) and oligogenic traits by means of forward genetic approaches, such as QTL analysis and genome-wide association study (GWAS), using experimental crosses (Uffelmann et al., 2021). If large effect QTLs are detected, we can assume that the target trait is monogenic or oligogenic. In such case, marker-assisted selection (MAS), in which breeders select broodstock according to the genotypes at the QTL, has economic advantages. However, it is often the case that large or even medium effect QTLs are not detectable for economic traits of plants, livestock, and aquaculture species as these are underpinned by large number of small-effect genes (i.e., polygenetic traits) (Crossa et al., 2017; Desta and Ortiz, 2014; Goddard and Hayes, 2009; Hosoya et al., 2017; Houston et al., 2020; Xu et al., 2020). In this case, GS has better selection response rather than MAS (Arruda et al., 2016).

Heritability and genetic architecture affect accuracy of genomic prediction (GP) of genomic estimated breeding values (GEBVs). Genomic prediction is the ranking process in GS for the selection of broodfish, and high accuracy of GP suggests high efficiency of GS breeding program (Meuwissen et al., 2001). In the process of GP, breeder creates a training group from a base population and then uses the phenotypes and genotypes of individuals from the group to train a prediction (regression) model. Next, the prediction model is applied to the selection candidates from the base population by substituting genotypes to estimate GEBV, which is a measure of the genetic potential for the trait of the candidate. Choose of a prediction model (i.e., linear mixed models, Bayesian models, machine learning models and deep learning models) also affects performance of GP as each model assumes different genetic architecture (Azodi et al., 2019) (detailed information is referred to in Section 1.2). Therefore, it is important to select the optimal model for each trait to achieve better genetic gain.

As described in Introduction, both heterobothriosis resistance and body size are important production traits for aquaculture of the tiger pufferfish. However, selecting one quantitative trait may improve or diminish others due to the genetic pleiotropy and/or linkage disequilibrium (Lynch and Walsh, 1998). For example, the breeding program which improves the resistance to sea lice possibly diminishes growth-related traits in farmed Atlantic salmon (Gjerde et al., 2011). Likewise, improving resistance to *H. okamotoi* may negatively affect growth-related traits in the tiger pufferfish. Thus, simultaneous genetic improvement of disease resistance and body size is most desirable, although it is complicated by the antagonistic genetic correlation. An index score, namely linear genomic selection index (LGSI), is highly expected to solve this problem (Ceron-Rojas et al., 2015). LGSI is calculated by a linear combination of GEBVs and corresponding weights (i.e., importance of the trait). Therefore, selection based on LGSI can take the importance of each trait into concern, rather than only one of the traits. In this chapter, I have examined the availability of LGSI for simultaneous genetic improvement of the two traits using simulation data.

1.1 Genetic dissection of heterobothriosis resistance and body size traits

To investigate the heritability and the genetic architectures of heterobothriosis resistance and standard length (SL), I have raised a test population derived from wild parents. These fish were subjected to an artificial infection trial to collect phenotypes and genotyped for genome-wide SNPs by means of target amplicon sequencing. These phenotype and genotype information were applied to genetic parameter estimation. Genome-wide association studies was also performed to clarify the genetic architecture of these traits.

Materials and methods

Tested population and artificial infection

The empirical experiments were performed in the Fisheries Laboratory, University of Tokyo (Hamamatsu, Shizuoka Prefecture, Japan). All samples (n = 240) were generated by a full-factorial mating among 10 wild males and 11 wild females, which are commercially caught from Wakasa Bay (Fukui Prefecture, Japan). For the mating, artificial fertilization was applied following the previous study (Kim et al., 2019) with minor modification. In brief, females were anesthetized with 200 mg/l of 2-phenoxyethanol and then ripened by injection of 150 μg/kg of luteinizing hormone-releasing hormone (LHRH, Sigma-Aldrich, St. Louis, MP, USA). Gametes were stripped from each individual and fertilized per male-female pair (110 pairs in total). Rearing and feeding conditions were set as previously described in Hosoya et al., 2014. In brief, fertilized eggs of each maternal half-sib family were mixed and kept in a hatching jar. After hatching, each maternal half-sib was kept in a one-ton tank for one month and then all families were mixed and cultured in a three-ton communal tank. All tanks were supplied with flow-through water and aeration. Fish larvae were fed nutrient-enriched live S-type rotifers, nutrient-enriched *Artemia* nauplii, and commercial pellets according to the developmental stage. At four months old, 240 fish were randomly collected and subjected to an artificial infection test.

To collect the phenotypes of tested fish, artificial infection was done following previous studies (Chigasaki et al., 2000; Kim et al., 2019). A day before the infection, fish were equally distributed into three identical one-ton experimental tanks (80 individuals/tank) supplied with *H. okamotoi*-free fresh seawater (UV treated and filtered). Meanwhile, eggs of *H. okamotoi* were collected from tanks containing infected fish and kept in a glass jar containing fresh seawater until infection. Hatching was induced by physical stimulation (shaking at 140 rpm for 10 min) and the density of oncomiracidia, the free-living larvae of *H. okamotoi*, in the suspensions was determined under the microscope just before the infection. At infection, the water depth of experimental tanks was adjusted to 15 cm, and approximately 4,000 oncomiracidia was introduced into each tank. At 3 h post-exposure, fish were transferred into three newly-setup one-ton holding tanks and reared for 32 days, when *H. okamotoi* reaches maturation and moves to the branchial cavity walls (BCW) (Ogawa, 2016). At the 32-day mark, fish were euthanized, measured for SL and the BCWs dissected from both sides. For each fish, the caudal fin was clipped and kept in 600 μl TNE8U buffer (Asahida et al., 1996) (10 mM Tris-HCl (pH7.5), 125 mM NaCl, 10 mM EDTA, 1% SDS, 8M urea) at room temperature to extract genomic DNA for genotyping. Collected BCWs tissues were kept in 10% formalin until counting the number of parasites under the microscope.

The parasites attached to the whole BCWs were counted under the stereoscopic microscope. The host resistance against *H. okamotoi* is assessed by parasite count on the whole BCWs (HC) (Supplementary Table S1-1).

Genotyping

I applied AmpliSeq technology (Sato et al., 2019), which uses polymerase chain reaction (PCR) to amplify the targeted regions for next generation sequencing. This approach has high repeatability of data and more robustness against *de novo* SNPs since only targeted regions can be consistently amplified by PCR (Sato et al., 2019). Specifically, genomic DNA was extracted using a Gentra Puregene Tissue Kit (QIAGEN, Hilden, Germany) following the manufacture's instruction and applied for AmpliSeq genotyping as previously described (Sato et al., 2019). In short, 3,187 genome-wide target regions were amplified by the first PCR with the custom AmpliSeq primer pools. After the adapter ligation and purification steps, PCR products were barcoded by a second PCR with 8-base index oligo sequences (Supplementary Table S1-1) for individual identification. The libraries of 240 individuals were pooled and sequenced on Illumina MiSeq System using the MiSeq reagent kit v2 (300 cycles) from both ends. The raw FASTQ reads were quality-trimmed using trimmomatic-0.36 (Bolger et al., 2014) with the following parameters: *ILLUMINACLIP TruSeq3-PE-2. fa:2:30:10, LEADING:19, TRAILING:19, CROP:146, HEADCROP:5, SLIDINGWINDOW:30:20, AVGQUAL:20, and MINLEN:60*. Then, trimmed reads were mapped onto the target regions of the reference genome, FUGU5/fr3 (Kai et al., 2011) using BWA-MEM (vv0.7.12) (Li, 2013). Reads with mapping quality values (MAPQ) less than 10 were removed by samtools (v1.7-2) (Li et al., 2009). Genotype calling of each sample was done using GATK-4.1.6.0 (Poplin et al., 2017) *HaplotypeCaller* with the following options: *--output-mode EMIT_ALL_CONFIDENT_SITES -ERC GVCF --stand-call-conf 30*. Obtained gVCF files were combined using GATK *CombineGVCFs* and then joint genotyping was performed using GATK *GenotypeGVCFs*. Variant filtering was done using vcftools (v0.1.5) (Danecek et al., 2011) with the following parameters: *--min-meanDP 15 --max-meanDP 500 --max-missing 0.7 --hwe 0.05 --minDP 10 --remove-indels*. The missing values of genotypes were imputed by LinkImputeR-1.2.1 (Money et al., 2017). First, the accuracy of 2 filters (i.e., the maximum missingness allowed per SNP and sample of 0.9 and 0.95) were tested with the parameters: *numbermaseked = 500*. The default setting was used for the other parameters. Subsequently, missing genotypes were imputed with a better filter (0.9). At first, SNPs which did not fulfill the maximum missingness per SNP and sample of 0.9 were filtered out, and then the missing genotypes were imputed. All samples were retained but 11 out of 6,718 SNPs were discarded. Subsequently, the imputation accuracy was accessed with *numbermasked* option (set as 500). PLINK (v1.0.7) (Purcell et al., 2007) *–recodeA* option was used to generate the allele coding matrix from the imputed VCF files. Command line scripts are supplied as Appendix.

Population structure

Population stratification gives biases to the results of heritability estimations, genome-wide association study (GWAS), and GP (Dandine-Roulland et al., 2016; Liu et al., 2015; Price et al., 2010). This is because the

regression model gives higher weight to the family-specific SNPs rather than SNPs in tight linkage disequilibrium with causative genes when stratification exists. Population structure can be easily grasped by visualization through dimensionality reduction methods, converting high-dimensional genomic data into low-dimensional maps without losing significant structure of the high-dimensional data. Traditionally, it is done by Principal Component Analysis, which reduces high-dimensional dimensions into a few to several principal components that explain the main patterns (Reich et al., 2008). Recently, a nonlinear dimension reduction technique, namely *t*-distributed stochastic neighbor embedding (*t*-SNE), gradually becomes popular in single-cell RNA-seq data analysis (Amir et al., 2013; Van Der Maaten and Hinton, 2008), and also shows good result in genetic studies (Li et al., 2017).

In this study, *t*-SNE was used to visualize population structure of the specimen. First, *t*-SNE transforms the genomic data into conditional probabilities that represent pairwise similarity in the high-dimensional space. Then, transformed data were applied to a heavy-tailed Student's *t*-distribution that measures pairwise similarities of corresponding samples in the low-dimensional embedding space. Finally, *t*-SNE minimized the sum of the Kullback–Leibler divergence between those two probability distributions (Kullback–Leibler divergence is the measure of the difference between two probability distributions.). The *t*-SNE analysis was implemented in *sklearn.manifold.TSNE* class of Python/scikit-learn-0.20.3. The perplexity was set as 20 and default values were used for the other parameters. Command line scripts are supplied as Appendix.

Heritability and genetic correlation

Heritability and genetic correlation were calculated by a multivariate linear mixed model as follows:

$$\boldsymbol{y}_i = \boldsymbol{X}_i\boldsymbol{\beta}_i + \boldsymbol{Z}_i\boldsymbol{u}_i + \boldsymbol{e}_i,$$

where $\boldsymbol{y}_i$ is a vector of phenotypes for trait *i* (*i* = 1 for transformed HC since HC is not normally distributed and 2 for SL); $\boldsymbol{X}_i$ and $\boldsymbol{Z}_i$ are incidence matrices for fixed effects $\boldsymbol{\beta}_i$ and random effects $\boldsymbol{u}_i$, respectively. The model assumes the random effects ($\boldsymbol{u}$) follow a multivariate normal distribution as $\boldsymbol{u} = [\boldsymbol{u}_1'\boldsymbol{u}_2']' \sim MVN(0, \boldsymbol{G} \otimes \boldsymbol{A})$ and the residuals ($\boldsymbol{e}$) follow $\boldsymbol{e} = [\boldsymbol{e}_1'\boldsymbol{e}_2']' \sim MVN(0, \boldsymbol{R} \otimes \boldsymbol{I})$; where $\boldsymbol{G}$ and $\boldsymbol{R}$ are the variance-covariance matrices of random effects and residuals for the two traits, respectively; $\boldsymbol{A}$ is the additive genetic relationship matrix constructed by *A.mat* function in R/rrBLUP-4.6 (Endelman, 2011) with the default settings; $\boldsymbol{I}$ is the identity matrix; $\otimes$ means the operation of Kronecker product. The model was solved by *mmer* function in R/sommer-4.0.1(Covarrubias-Pazaran, 2018, 2016) to solve the equations. The heritability (h_i^2) was computed as:

$$h_i^2 = \frac{\sigma_{g_i}^2}{\sigma_{g_i}^2 + \sigma_{e_i}^2},$$

where $\sigma_{g_i}^2$ and $\sigma_{e_i}^2$ are the genetic variance and the residual variance for trait *i*, respectively. Then, the genetic correlation (r_g) was computed as:

$$r_g = \frac{\sigma_{g_1,g_2}}{\sqrt{\sigma_{g_1}^2\sigma_{g_2}^2}},$$

where σ_{g_1,g_2} is the genetic covariance between two traits.

The genetic correlation estimated by means of the multivariate model could be biased from the phenotypic correlation. Therefore, I further tested correlation between GEBV for each trait using GBLUP (genomic best linear unbiased prediction) model. In the prediction model for HC, SL was included as the covariate to minimize non-genetic effects from SL. The prediction models are described as following:

$$\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{Zu} + \boldsymbol{\varepsilon},$$

where $\boldsymbol{y}$ is a vector of phenotypes; $\boldsymbol{X}$ is an incidence matrix for the fixed effect $\boldsymbol{\beta}$ (for the prediction of HC, SL was added as a covariate); $\boldsymbol{Z}$ is an identity matrix for the random effects $\boldsymbol{u}$, which models the breeding values; $\boldsymbol{\varepsilon}$ is a vector of residuals. The normality was assumed for random effects ($\boldsymbol{u}$) and residuals ($\boldsymbol{\varepsilon}$) as $\boldsymbol{u} \sim N(0, \boldsymbol{K}\sigma_u^2)$ and $\boldsymbol{\varepsilon} \sim N(0, \boldsymbol{I}\sigma_\varepsilon^2)$, respectively; where $\boldsymbol{K}$ is a marker-based relationship matrix (Endelman, 2011); $\boldsymbol{I}$ is an identity matrix. $\boldsymbol{K}$ was constructed as described above; GEBVs were estimated by restricted maximum likelihood (REML) algorithm using *kin.blup* function in R/rrBLUP-4.6. Command line scripts are supplied as Appendix.

Genome-wide association study (GWAS)

To investigate the associated markers with transformed HC and SL, GWAS was performed based on the linear mixed model:

$$\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{Zg} + \boldsymbol{S\tau} + \boldsymbol{\varepsilon},$$

where $\boldsymbol{y}$ is the vector of the phenotypes; $\boldsymbol{\beta}$ is a vector of fixed effects other than SNP effects; $\boldsymbol{g}$ is the vector of random effects that models the polygene background; $\boldsymbol{\tau}$ is a vector of fixed effects which represent the additive SNP effects; $\boldsymbol{X}$, $\boldsymbol{Z}$, and $\boldsymbol{S}$ are incidence matrices relating to $\boldsymbol{\beta}$, $\boldsymbol{g}$, and $\boldsymbol{\tau}$, respectively. $\boldsymbol{\varepsilon}$ is a vector of normal residuals. $\boldsymbol{g}$ and $\boldsymbol{\varepsilon}$ follow the normal distributions as $\boldsymbol{g} \sim N\left(0, \boldsymbol{K}\sigma_g^2\right)$ and $\boldsymbol{\varepsilon} \sim N(0, \boldsymbol{I}\sigma_\varepsilon^2)$, respectively; where $\boldsymbol{K}$ is the realized relationship matrix described above. A restricted maximum likelihood (REML) algorithm was performed to solve the linear mixed model using *GWAS* function of R/rrBLUP-4.6 with the parameter: n.PC = 10. The *p*-values were calculated for each SNP marker. The Bonferroni-corrected significant threshold was set as $\alpha = 7.45 \times 10^{-6}$ (0.05 divided by the number of SNPs: 6,707).

To further examine the effects of SNP markers, association analysis assuming the following Bayes C model was performed:

$$\boldsymbol{y} = \mu\mathbf{1}_n + \boldsymbol{X\beta} + \sum_{i=1}^{p} \boldsymbol{z}_i\boldsymbol{g}_i + \boldsymbol{\varepsilon},$$

where $\boldsymbol{y}$, $\boldsymbol{X}$, $\boldsymbol{\beta}$ and $\boldsymbol{\varepsilon}$ are same as GBLUP; μ is an intercept; $\mathbf{1}_n$ is a vector of one; p is the total number of SNP loci for each individual; $\boldsymbol{z}_i$ is a scalar of genotypes at SNP i; $\boldsymbol{g}_i$ is a scalar of random effects that represent the genetic effects for SNP i following a mixture of scaled-t distribution and a point of mass at zero. The model was

solved using R/BGLR-1.0.8 (Pérez and De Los Campos, 2014) with nIter = 10,000 and burnIn = 2,000. Command line scripts are supplied as Appendix.

Results

Tested population and artificial infection

Specimens (n = 240) produced by artificially crossing 11 wild males and 10 wild females were subjected to an artificial infection for 37 days at four months old. The phenotypic mean was 15.85 (± 9.15 S.D.) for HC and 9.83 (± 0.78 S.D.) for SL (Figure 1-1 and Supplementary Table S1-1). As the plot shows, the distribution of HC was non-normal (Shapiro–Wilk test: $p = 3.79\times10^{-6}$, alpha level = 0.05) while SL approximated a normal distribution (Shapiro–Wilk test: p = 0.406, alpha level = 0.05). Therefore, I applied a square-root transformation on (HC +1), approximating a normal distribution (Shapiro–Wilk test: p = 0.235, alpha level = 0.05). Transformed HC was used in the following genetic analysis. Weak but significant phenotypic correlation was observed between HC and SL (Pearson's r analysis: r = 0.157, p = 0.015; 95% confidence interval: $0.031 \leq r \leq 0.278$).

Genotyping

The MiSeq sequencing generated an average of 174,870 (± 83,576 S.D.) raw reads per fish. Amplicon sequence reads have been deposited in the DDBJ Sequence Read Archive (DRA Accession: DRA010341) (Supplementary table S1-2). After the quality-trimming step, the mean number of reads for each fish was 161,426 (± 83,576 S.D.) with the mean read length of 124 bp. The survived reads were mapped onto a reference fugu genome (FUGU5/fr3) for SNP calling. Following the quality filtration of SNPs, 6,718 putative SNPs were yielded. Missing SNPs were imputed using LinkImputeR (Money et al., 2017). At this imputation step, 11 SNPs were discarded, and 6,707 imputed SNPs were called for each individual with the imputation accuracy of 0.888.

Population structure

Population structure, which can bias the genetic parameter estimation, was examined by t-SNE analysis (Van Der Maaten and Hinton, 2008) based on SNP data. As seen in Figure 1-2, I did not observe clear clusters or strong stratification within the tested samples. The distribution of HC showed weak linear relationship between both x and y coordinates (Pearson's r = 0.056 and 0.210, respectively). This population structure ensured limited biases to the results of heritability estimations, GWAS, and GP.

Heritability and genetic correlation

To investigate the extent of genetic effects on the phenotypic variation, heritability was estimated by a multivariate linear mixed model. Moderate heritability was observed for each trait (transformed HC: h_2 = 0.308

± 0.123 S.E.; SL: h_2 = 0.405 ± 0.131). With the same model, the strength of the genetic correlation between the transformed HC and SL was also estimated. A moderate antagonistic genetic correlation (r_g = 0.228) was observed between the traits, where large individuals were suffering from higher parasitic loads. This genetic correlation could be, at least partly, due to the phenotypic correlation, although phenotypic correlation between HC and SL was weak as described above. Therefore, I tested correlation between GEBV for each trait using a univariate linear model (i.e., GBLUP); SL was included as the covariate in the prediction model for HC to minimize non-genetic effects from SL. If an antagonistic genetic correlation exists between the two triats, the GEBVs would also show a positive correlation. As the results, I found moderate but significant positive correlation (Pearson's $r = 0.252$, $p = 7.67 \times 10^{-5}$).

Genome-wide association study (GWAS)

GWAS was applied to detect loci highly associated with transformed HC and SL (Figure 1-3). None of these loci exceeded the significance threshold of 7.45 (= $-\log_{10}$ (0.05/6,707)). Bayes C model supported the results that each SNP has minimal effects (effect absolute value < 0.1) and the two traits are polygenic.

1.2 Model comparison of genomic prediction

In the previous section, moderate heritability was observed for both heterobothriosis resistance and SL (transformed HC: $h^2 = 0.308 \pm 0.123$ S.E.; SL: $h^2 = 0.405 \pm 0.131$). In addition, the polygenetic nature of these traits was confirmed and high potential for genetic improvement by GS was indicated. In GS breeding program, genetic gain for a target trait depends on the accuracy of GP. Thus, many advanced regression models have been proposed to achieve higher accuracy of GP. For instance, a linear mixed model, GBLUP uses the marker-based realized relationship matrix and the best linear unbiased prediction (BLUP) to estimate the GEBVs. The Bayesian models (e.g., Bayes A, B, C, Ridge, and LASSO) assume different priors to manipulate the variance of genetic values under Bayesian rules and use stochastic methods, namely Markov chain Monte Carlo (MCMC), to solve the linear mixed model (Gianola et al., 2013). While these approaches assume parametric additive models, non- (or semi-) parametric model, such as a reproducing kernel Hilbert space (RKHS) regression, can model multiple and complex interactions among loci, potentially arising over whole genome with nonparametric treatments by introducing smoothing parameters as variance components (Gianola et al 2006; Gianola and van Kaam, 2008). On the other hand, deep learning models were developed under the inspiration of information processing processes in the biological nervous system. These models are built from lots of non-linear sub-models (neurons) connected by the neuron circuits (mimicking how neurons are connected each other) to capture non-linear interactions between the phenotypes and genotypes by feature extraction and transformation (Pérez-Enciso and Zingaretti, 2019). The potential of each prediction model depends on the genetic architecture of the trait and should be tested per trait. In this section, model comparison of GP among 12 models was done for the two traits. Since prediction accuracy was not available for non-linear models, predictive ability was used as the evaluation metric for model comparison among all models

Materials and methods

Predictive abilities of GP

Predictive abilities were obtained under 12 regression models described below. The tenfold cross-validation scheme was applied for predictive ability and accuracy calculation following the procedure described in Hosoya et al. (2018). Samples were randomly and equally divided into ten subsets: one for testing and the remaining for training. The phenotypic values of the test set were masked, and the regression model was trained using the training set. GEBVs of the test set were predicted and predictive ability was calculated as the correlation (Pearson's *r*) between GEBVs and observed values of the test set. Then, prediction accuracy was calculated for the GBLUP and Bayesian models as predictive ability divided by the square root of heritability, which was calculated as described previously. This step was repeated with rotating the test sets among the five subsets, and the average of Pearson's *r* was obtained. This cross-validation process was repeated 10 times to obtain the mean and the standard error of the mean (S.E.) for the predictive ability and accuracy. Transformed HC instead of the original phenotype was used in GBLUP and Bayesian models. To generate the consistent random state for sampling among 12 models, I fixed seeds for random sampling among the models.

Genomic best linear unbiased prediction (GBLUP)

GBLUP was implemented as Chapter 1, Section 1.

Bayesian models

The models of Bayes A, B, C, Ridge, and LASSO (Habier et al., 2011; Meuwissen et al., 2001; Park and Casella, 2008) are expressed as follows:

$$\boldsymbol{y} = \mu\mathbf{1}_n + \boldsymbol{X\beta} + \sum_{i=1}^{p} \boldsymbol{z}_i\boldsymbol{g}_i + \boldsymbol{\varepsilon},$$

where $\boldsymbol{y}$, $\boldsymbol{X}$, $\boldsymbol{\beta}$ and $\boldsymbol{\varepsilon}$ are same as GBLUP; μ is an intercept; $\mathbf{1}_n$ is a vector of one; p is the total number of genotypes for one individual; $\boldsymbol{z}_i$ is a vector of genotypes at SNP *i*; $\boldsymbol{g}_i$ is a vector of random effects that represent the genetic effects for SNP *i*. following a specific prior distribution. Bayes A assumes a scaled-*t* distribution for the prior while Bayes B assumes a mixture of gaussian distribution and a point mass at zero. The prior of Bayes C is a mixture of scaled-*t* distribution and a point of mass at zero. The prior of Bayes Ridge and Bayes LASSO is a normal distribution and a double exponential distribution, respectively. These models were solved using R/BGLR-1.0.8 (Pérez and De Los Campos, 2014) with nIter = 10,000 and burnIn = 2,000.

Bayesian reproducing kernel Hilbert spaces regressions (Bayesian RKHS)

Bayesian RKHS is a Bayesian approach of semi-parametric regression (De Los Campos et al., 2010) structured as:

$$\boldsymbol{y} = \mu\mathbf{1}_n + \boldsymbol{u} + \boldsymbol{\varepsilon},$$

where each parameter is the same as the Bayesian models, while $\boldsymbol{u}$ and $\boldsymbol{\varepsilon}$ follow the normal distribution as $\boldsymbol{u} \sim N(0,\ \boldsymbol{K}\sigma_u^2)$ and $\boldsymbol{\varepsilon} \sim N(0,\ \boldsymbol{I}\sigma_\varepsilon^2)$, respectively; where $\boldsymbol{K}$ is a Gaussian reproducing kernel (bandwidth parameter = 1) which approximates the marker-based relationship matrix and $\boldsymbol{I}$ is an identity matrix. The model was solved using R/BGLR-1.0.8 with nIter = 10,000 and burnIn = 2,000.

Support vector machine regression (SVR)

SVR method can be viewed as a convex optimization problem that finds a function from observed values to estimated values at most ε-deviation from all observed values while balancing the model complexity and prediction error (Awad et al., 2015; Vapnik, 1995). The method of Lagrange multipliers is used to solve the optimization problem, and the derived approximate function follows:

$$f(\boldsymbol{x}) = \sum_{i=1}^{N} (a_i^* - a_i) k(\boldsymbol{x}, \boldsymbol{x}_i) + b,$$

where the input $\boldsymbol{x}$ is a vector of all genotypes for a single sample; N is the sample size; a_i^* and a_i are Lagrange multipliers; $k(\boldsymbol{x}, \boldsymbol{x}_i)$ is a kernel function; $\boldsymbol{x}_i$ is a vector of genotypes for individual i; b is a residual. SVR-linear, SVR-poly, and SVR-rbf using linear, polynomial, and radial basis for kernel function, respectively. The SVR models were implemented by the *sklearn.svm.SVR* function in Python/scikit-learn-0.20.31 (Pedregosa et al., 2011). The *gamma* parameter was set to '*auto*' and the default setting was used for the other parameters.

Neural networks

Feedforward neural networks (FNN), inspired by the biological neural network, can model genotype-phenotype regression (Gianola et al., 2011). Neural cells were modeled by non-linear functions (or activation function) and the network was mimicked by the chain structure. My FNN had one input layer, two hidden layers, and a regression output. The number of input units was 6,707, equivalent to the number of SNPs. The first hidden layer has 200 hidden units and the second 20. The rectified linear unit (ReLU) was used as an activation function in hidden layers. FNN was trained by minimizing the loss function, that is, the mean squared error in this case:

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$$

where *n* is the sample size of the training group**;** y_i and $\hat{y}_i$ are observed value and the predicted value of individual i, respectively.

Multi-task deep neural network (Multi-task FNN) is based on an assumption where HC and SL share underlying genetic architecture to some extent. These models can improve the accuracy of estimation of the main output using the related auxiliary task as an inductive bias to the main task in reproducing kernel Hilbert space (Caruana, 1997; Widmer and Rätsch, 2011). The model has one input layer, two hidden layers, one main regression output, and one auxiliary regression output. The first hidden layer has 200 units, which is a sharing layer for both tasks. Both outputs have separated second hidden layers that differ in the number of the hidden units (20 for the main trait and 100 units for the auxiliary trait). For the estimation of HC, the main regression output is for HC and auxiliary regression output is for SL. The model setting of the main regression and auxiliary regression output was reversed for SL estimation. The activation function and the loss function were the same as the FNN model described above. FNN and multi-task FNN were implemented in Python/keras-2.4.3 package (Chollet and others, 2015) with tensorflow-gpu-2.2.1 backend (Abadi et al., 2016). FNN used "Adam" optimizer, and multi-task FNN used "RMSprop" optimizer, both with the default parameters. Both models were trained by *model.fit* method in Python/keras with the parameters as epochs = 30, batch_size = 128, and others followed the default. Many combinations of model architecture, loss function, activate the function, and optimizer was arbitrarily chosen and tested to find the models here that have a high accuracy of GP for HC.

Results

Predictive abilities for transformed HC ranged from 0.248 to 0.344 under 12 models (Table 1-1). Among them, SVR-poly and SVR-rbf models were inferior, while two deep learning models were slightly better. Predictive abilities for SL ranged from 0.340 to 0.481 under 12 models (Table 1-1). In contrast to the case of HC, the two SVR based models ranked at the top for prediction of SL, and deep learning models were inferior. Bayes RKHS and GBLUP models showed good performance in both traits. Moderate predictive ability of GP for both of the traits suggested the availability of GS since these results ensure the selection accuracy of broodfish.

1.3 Breeding strategy for simultaneous improvements of both heterobothriosis resistance and body size

In the previous sections, the availability of GS for both heterobothriosis resistance and body size was shown from the moderate heritability and predictive ability. However, the results also showed an antagonistic genetic correlation between these traits (r_g = 0.228); larger individuals are possibly suffering from higher parasitic loads. This undesired correlation complicates simultaneous genetic improvements of the two traits. One of the conventional methods for multiple-trait improvement is the linear selection index (LSI) method developed by Smith and Hazel (Hazel, 1943; Smith, 1936). Net genetic merit (i.e., LSI) of each animal is calculated from each target trait and used for ranking breeding candidates. To maximize the selection response, a general LSI is computed by a linear combination of phenotypes or EBVs (estimated breeding values) and the corresponding coefficients. Extensive LSI methods have been proposed (Cerón-Rojas and Crossa, 2018), as determined by the method of coefficient calculation. For instance, the desired gain selection index allows breeders to restrict traits according to the expected change of genetic gain values of traits (Itoh and Yamada, 1987). In the era of GS, those LSI methods can be directly applied to compute the linear genomic selection index (LGSI), which showed higher efficiency in both simulation and real data, compared to pedigree-based LSI (Ceron-Rojas et al., 2015). Although LGSI showed great advantages, successful applications of LGSI still largely depend on the accurate estimation of GEBVs and genetic parameters (Togashi et al., 2011), which are sensitive to many factors, including the genetic architecture of target traits, population structure, genotyping technologies, etc. (Daetwyler et al., 2010; Guo et al., 2014; Solberg et al., 2008). Consequently, an LGSI method might have different performances in different cases. Therefore, it is essential to find the optimal strategy incorporating LGSI and examine its performance in each breeding program. The GS breeding simulator will be a practical tool that approximates the real genetic progress by sophisticated modeling of the meiosis and GS procedure at the DNA level (Daetwyler et al., 2013). Further, as regards selection targeting disease resistance traits in aquaculture, a recent simulation study of acute hepatopancreatic necrosis disease (AHPND) in shrimp (*Litopenaeus vannamei*) showed that GS was superior to pedigree-based methods (Wang et al., 2019). Therefore, with the assistance of simulation, the breeding strategies incorporating LGSI are expected to greatly accelerate the simultaneous genetic improvement of disease resistance and growth-related traits.

Materials and methods

Simulation

To investigate the breeding strategy that can improve SL and HC simultaneously, six scenarios different in recurrent selection schemes were simulated for ten generations with 50 replicates using R/AlphaSimR-0.11.0 package (Faux et al., 2016). The tested scenarios were named for the selection scheme applied: random mating (RAND), GS on HC only (GS_{HC}), GS on SL only (GS_{SL}), Smith-Hazel index ($S1_{SHI}$ and $S2_{SHI}$), and Desired gains index (S_{DGI}). The workflow of the simulation study is illustrated in Figure 1-4. In short, RAND was based on random mating while GS_{HC} and GS_{SL} were based on GS on either of the traits. GEBV was estimated by GBLUP. In $S1_{SHI}$, selection candidates were ranked based on the Smith-Hazel index. Since economic importance for each trait has not been evaluated in the tiger pufferfish aquaculture industry, I assume both traits have equal economic weights, which is w = [-1, 1] for HC and SL for $S1_{SHI}$ (HC is expected to

decrease by selection). For S_{DGI}, *d* was set as [-3, 0.3] for HC and SL, so that SL can be improved preferentially while HC can be reduced by 30% after 10 generations (-3*10/100 = -30%). To compare the two selection index methods, I also run an additional scenario ($S2_{SHI}$) based on Smith-Hazel index, where the economic weight for each trait was set as same as the designed weights of S_{DGI} (*w* = [-3, 0.3]).

The simulated population was generated by *runMaCS2* function in R/AlphaSimR package (Gaynor et al., 2020). First, all scenarios began with a founder population of 10,000 individuals was simulated assuming: the effective population size of 1,000, mutation rate of 2.5×10^{-8}, no inbreeding in founder individuals. The ploidy (*n* = 2), the number of chromosomes (*n* = 22), and genetic and physical size of each chromosome (Morgans and base pairs, respectively) were set according to the reference genome sequence integrated with the genetic map of the tiger pufferfish, FUGU5/fr3 (Kai et al., 2011). The relative ratio of recombination in females compared to males was set as 1.82 according to FUGU5/fr3. The phenotypic mean of SL was set in accordance with the phenotyping result and that of HC was set as 100 to avoid minus values of phenotypes after genetic improvement. Phenotypic variance, genetic variance, heritability, and genetic correlation were simulated referring to the analysis result using empirical data obtained in this study. The gender of each individual was randomly assigned. For each trait, 500 QTLs were placed per chromosome. The number of SNP markers per chromosome was equal to that detected in the Ampliseq custom panel (*n* = 6,707 in total) and randomly placed over each chromosome to form an SNP chip *in silico*. To initiate the GS breeding program, 20 sires and 20 dams were randomly sampled from the founder population to perform a full-factorial mating, and then each parent pair generates 20 progenies and in total 8,000 progenies were produced. From the progeny pool, 2,000 fish were randomly picked up as the broodstock population (F_0). The relatively small number of parents were used in this simulation study compared with the practical breeding programs due to the limited computer resources, but it will be enough for a test study.

Subsequently, the recurrent selection schemes were performed for ten generations with 50 replicates independently among six scenarios (RAND, GS_{HC}, GS_{SL}, $S1_{SHI}$, $S2_{SHI}$, and S_{DGI}). In each generation, according to the scenario-specific criteria, 20 sires and 20 dams were selected and crossed with a full-factorial mating system to create next-generation where each mating cross generated 20 progenies (total 8,000 progenies). Only 2,000 fish out of 8,000 progenies remained as the broodstock candidates for the next generation. This process was performed for a total of ten generations with 50 replicates. The broodstock population produced in the i-th generation was noted as F_i (i = 1,2,3…10). The scenario-specific criteria were as following. In the RAND scenario, parental individuals were randomly selected from the candidates (n = 2,000) in each generation. The individuals with high GEBVs for only a single trait were chosen in the GS_{HC} and GS_{SL} scenario, while the ones with high LGSIs in $S1_{SHI}$, $S2_{SHI}$ and S_{DGI} Scenario. For each generation, in GS_{SL} scenario, GBLUP model was trained using all candidates (n = 2,000) and broodfish were directly selected from these individuals, whilst, in GS_{HC}, S_{SHI} and S_{DGI} scenario, GBLUP model was trained using half of the candidates (n = 1,000) and broodfish were selected from the remaining (n = 1,000) since fish should be sacrificed to obtain HC phenotype. The GBLUP model was implemented by *RRBLUP* function in R/AlphaSimR package.

Linear genomic selection index

LGSI for $S1_{SHI}$, $S2_{SHI}$ and S_{DGI} was constructed as:

$$\boldsymbol{LGSI} = \boldsymbol{b}'\widehat{\boldsymbol{y}},$$

where $\boldsymbol{b}$ is a vector of index coefficients; $\widehat{\boldsymbol{y}}$ is a vector of GEBVs. $\boldsymbol{b}$ for $S1_{SHI}$ and $S2_{SHI}$ was computed as:

$$\boldsymbol{b} = \boldsymbol{P}^{-1}\boldsymbol{A}\boldsymbol{w},$$

where $\boldsymbol{P}$ and $\boldsymbol{A}$ are phenotypic and genetic variance-covariance matrices, respectively; $\boldsymbol{w}$ is the economic importance of both traits and set as [-1, 1] assuming both traits have equal economic weights (HC is expected to be decreased by selection) for $S1_{SHI}$. $\boldsymbol{P}$ was obtained by *varP* function of R/AlphaSimR, and $\boldsymbol{A}$ was obtained by *mmer* function of R/sommer-4.0.1 following the same procedure for the estimation of genetic correlation except that original HC value was used. On the other hand, $\boldsymbol{b}$ for S_{DGI} was computed as:

$$\boldsymbol{b} = \boldsymbol{P}^{-1}\boldsymbol{A}(\boldsymbol{A}\boldsymbol{P}^{-1}\boldsymbol{A})^{-1}\boldsymbol{d},$$

where $\boldsymbol{P}$ and $\boldsymbol{A}$ are the same as S_{SHI} while $\boldsymbol{d}$ is a vector of desired gains. The combination of $\boldsymbol{d}$ can be chosen arbitrary depending on the breeding goal of the program. In this study, I set $\boldsymbol{d}$ at [-3, 0.3] for HC and SL so that SL can be improved preferentially while HC can be reduced by 30% after 10 generations. To compare between different selection index methods, the vector of economic importance, [-3, 0.3], which is the same with $\boldsymbol{d}$ in S_{DGI} was assigned to $\boldsymbol{w}$ in $S2_{SHI}$.

Results

The availability of LGSI methods for simultaneous improvements of HC and SL was tested using simulation data. The scenario with only random selection, RAND, showed that true breeding values (TBVs) at F_{10} generation fluctuated up and down around the TBVs at F_0 for both of the traits, and no obvious genetic changes were observed (Figure 1-5. **a**). GS_{HC} and $GS_{SL,}$ the scenarios which performed selection on single trait, TBVs of the targeted trait was improved while hindered in the other (Figure 1-5. **b** and **c**). The scenario ($S1_{SHI}$ and $S2_{SHI}$) applying Smith-Hazel index showed average TBV for both SL and HC decreased, i.e., smaller fish with less parasite loads were selected (Figure 1-5. **d** and **e**). As expected, only S_{DGI} could improve the two traits simultaneously, where true breeding values (TBVs) of parasite load (HC) decreased while SL increased in each generation (Figure 1-5. **f**).

Discussion

In this chapter, the possibility of GS for heterobothriosis resistance and SL of the tiger pufferfish was tested from empirical data. In addition, a GS breeding strategy that can improve the resistance trait concurrently with SL was designed using a simulation study.

With 6,707 SNP makers, moderate estimated heritability of transformed HC (h^2 = 0.308, SE = 0.123) and SL (h^2 = 0.405, SE = 0.131) were obtained, indicating selective breeding for those traits is feasible. The estimated heritability was comparable to those estimated for resistance against sea lice in Atlantic salmon (h^2 = 0.22 to 0.33 with 35k SNPs) (Tsai et al., 2016) and bacterial cold water disease resistance (survival days) in farmed rainbow trout (h^2 = 0.33 with 35k SNPs) (Vallejo et al., 2017). Although the small SNP panel can capture the moderate heritability for HC and SL in the tiger pufferfish, the relatively large standard error suggested that a middle- or large-scale study was still needed to confirm the generality of this results. In this study, significant SNPs were not detected in GWAS. It has been shown that even with the small SNP panel and small sample size, strong effect SNPs (the sex-determining SNP) can be detected in a cultured population of the tiger pufferfish (Sato et al., 2019). Therefore, GWAS result suggests the parasitic resistance is controlled by a large number of quantitative trait locus (QTL) with small or moderate effects, and marker-assisted selection is not feasible, although larger sample size and more SNPs may increase the credibility of the GWAS result as same as the heritability estimation.

The predictive abilities for HC estimated under 12 models were moderate (0.248–0.344), and within the range observed for other disease resistance traits examined in other fish species (Odegård et al., 2014; Palaiokostas et al., 2018, 2016; Robledo et al., 2018a). The predictive abilities of Bayesian hierarchical linear models (i.e. Bayes A, B, C, LASSO, and Ridge) were similar (0.303–0.312) and scarcely higher than the GBLUP model (0.307 ± 0.018) for HC. This suggests that these linear models did not greatly differ regarding the predictive ability and the assumptions of the prior distribution of genetic effects have a limited impact on this trait. Bayes RKHS showed slightly better performance in HC compared to these linear models. For SVR-poly and SVR-rbf models, relatively low abilities for HC were observed, however, high abilities were found for SL. Since the default hyperparameters were used in the SVR models, hyperparameter tuning may aid achievement of better performance for HC as in the case of the previous study (Azodi et al., 2019). The architectures of FNN and multi-task FNN were tuned to achieve high predictive ability of GS for HC, however, the same architecture was applied to calculate the predictive ability of GS for SL. As expected, these models resulted in high predictive ability for HC but low for SL. This indicates that a deep learning model is task-specific and high accuracy can be obtained with careful optimization as described previously (Pérez-Enciso and Zingaretti, 2019). However, a great improvement in predictive ability was not achieved by FNN methods compared to GBLUP and Bayesian models even with the model complexity.

The simulation study showed the availability of DGI for simultaneous genetic improvement in HC and SL even when the unfavorable antagonistic genetic correlation was assumed. The two scenarios incorporated the Smith-Hazel index showed the undesired consequences, where the average TBV for both SL and HC increased (smaller HC is favored). This happened because the breeding scheme only selected the individuals with the top LGSI values, but the high LGSI calculated by the Smith-Hazel index method does not guarantee the selected

individuals are superior in both of the traits (Kempthorne and Nordskog, 1959), especially when target traits show a negative correlation. On the other hand, DGI, a variation of the selection index methods, allows selection with restrictions on multiple traits via the desired gains vector (*d*). In this study, the *d* vector was set with aiming to reduce HC by 30% during 10 generations while maximizing SL. The desired gains vector (*d*) can be further optimized by comparing simulation scenarios with various *d* to achieve the self-defined breeding goal. Unfavorable genetic correlation between body size and disease resistance is commonly observed in aquaculture species, e.g. vibriosis in Atlantic cod (Bangera et al., 2011), bacterial cold water disease in rainbow trout (Evenhuis et al., 2015), and piscirickettsiosis in coho salmon (Bangera et al., 2011; Evenhuis et al., 2015; Yáñez et al., 2016). Therefore, it is expected that DGI or the similar LGSI method can be widely applied for the simultaneous improvement of disease resistance trait and growth-related traits, which are the primary targets of most breeding programs.

In summary, the availability of GS for HC and SL in the tiger pufferfish under a standard diet was confirmed in this small-scale study. Moderate heritability for both traits suggest the genetic return from GS is high. GBLUP and Bayesian linear regression models showed similar prediction performance for these traits. Although an unfavorable antagonistic genetic correlation was suggested between the two traits, the GS breeding strategy incorporating DGI can be a solution for the simultaneous genetic improvement.

Chapter 2. Genomic prediction for heterobothriosis resistance and body size traits in a middle-scale population

The availability of genomic prediction (GP) for heterobothriosis and body size under a standard diet was confirmed in Chapter 1 using a small number of samples (n = 240) produced from wild tiger pufferfish with a low-density SNP panel (6,707 SNPs). This low-cost, small-scale experiment worked sufficiently as a pilot study. However, such small sample size and low-density SNP panel limit the credibility of genetic parameters (i.e., large standard error of heritability: HC: h^2 = 0.308 ± 0.123 S.E.; SL: h^2 = 0.405 ± 0.131) and the accuracy of GP (Karaman et al., 2016; Kriaridou et al., 2020; Uffelmann et al., 2021). In this chapter, I have examined the generality of the results obtained from the small-scale experiment using larger number of reference fish with higher density SNP set. Specifically, I used 1,100 four-month-old individuals produced from commercially cultured fish and a medium density SNP set (12,548 imputed SNPs per fish) to examine the genetic basis of heterobothriosis resistance and body size. Although the use of low-density SNP panels can inferior the credibility of genetic parameters and the accuracy of GP, it can reduce the genotyping costs and improve cost efficiency in total when the accuracy is comparable to that estimated with a high-density panel (Kriaridou et al., 2020; Tsairidou et al., 2020). To search test the minimum SNP density which do not have strong negative impacts on predictive ability for the two traits, I also examined the effects of the number of SNPs on GP under GBLUP model.

2.1 Genetic dissection of heterobothriosis resistance and body size traits

Materials and methods

Test population

Fish culture and the experimental infection trial were performed at Nagasaki Prefectural Institute of Fisheries (NPIF, Nagasaki, Japan). Test fish were produced by means of artificial fertilization following the previous studies (Miyaki et al., 1998; Yoshikawa et al., 2020). A full-factorial mating among were crossed in a full-factorial manner. Specifically, gametes were stripped from ten males and four females, consisting of commercially raised fish derived from families cultured for several generation, and crossed in a full-factorial manner per male–female pair (40 pairs). Then, eggs from each full-sib family were incubated in one-ton tanks. Hatchlings from each full-sib family were transferred and reared communally. Rearing conditions follows the previous studies (Miyaki et al., 1998; Yoshikawa et al., 2020). Fish were fed with nutrient-enriched live L-type rotifers, *Artemia* nauplii and commercial pellets based on their developmental stage. Ultraviolet-sterilized seawater and aeration was provided to all tanks. At four-month-old, 1,200 fish were randomly collected as the specimens for the artificial infection.

Artificial infection

The phenotypes for host resistance against *H. okamotoi* was collected using the conventional artificial infection following Chapter 1, section 1 with modification. Specifically, four identical two-ton experimental

tanks with *H. okamotoi*-free fresh seawater were prepared for total 1,200 tested fish (300 fish/tank). After adjustment of the water depth of experimental tanks (15 cm), approximately 7,000 oncomiracidia were introduced into each tank (around 23 oncomiracidium per fish) for 3-hour exposure. Subsequently, fish were transferred into eight, newly-setup one-ton tanks (150 fish/tank) and reared until *H. okamotoi* reaches maturation (46 days). At 46-days post infection, SL of total 1,106 fish were recorded; meanwhile, branchial cavity walls of each fish were dissected and kept in 10% formalin until counting matured *H. okamotoi* (HC) under a stereo microscope.

Genotyping

For 1,100 individuals, genomic DNA was extracted from fin clips by Gentra Puregene Tissue Kit (QIAGEN, Hilden, Germany) following the manufacture's instruction. These samples were genotyped using a nontargeting PCR-based genotyping by sequencing technology, i.e., genotyping by random amplicon sequencing, direct (GRAS-Di) (Enoki, 2019; Enoki and Takeuchi, 2018) as described in the previous literatures (Hosoya et al., 2019). In brief, DNA template was amplified by the first PCR with the primer pools included Illumina Nextera adaptor sequences plus 3-base random oligomers. Then, the diluted PCR products were barcoded by a second PCR with Illumina multiplexing 8-base dual index and P7/P5 adapter sequence. The libraries were prepared using HiSeq X Ten Reagent Kit v2.5 and sequenced on Illumina Hiseq X system (Illumina, San Diego, USA). Sequencing was done by Genebay, Inc (Kanagawa, Japan).

Raw reads were trimmed by trimmomatic (v0.39) (Bolger et al., 2014) with the parameters: ILLUMINACLIP: *TruSeq2-PE.fa:2:30:10 SLIDINGWINDOW:30:20 AVGQUAL:20*. Trimmed reads were mapped to the reference genome FUGU5 by BWA-MEM (v0.7.17-r1188) (Li, 2013) with adding RG tags for each reads. Then, reads with mapping quality values (MAPQ) less than 10, unmapped reads, reads with multiple primary alignments, and reads with supplementary alignments were removed by samtools (v1.9) (Li et al., 2009). Genotype calling of each sample was done using GATK (v4.1.6.0) (Poplin et al., 2017). *HaplotypeCaller* with the following options: *--output-mode EMIT_ALL_CONFIDENT_SITES -ERC GVCF –stand-call-conf 30*. Obtained gVCF files were combined using GATK *GenomicsDBImport* and then joint genotyping was performed using GATK *GenotypeGVCFs*. Variant filtering was done using vcftools (v0.1.5) with the following parameters: *--remove-indels –min-alleles 2 –max-alleles 2 –minQ 20 –min-meanDP 20 –minDP 10 –max-meanDP 300 –maf 0.01 –hwe 0.05 –max-missing 0.7*. Finally, the missing genotypes of the filtered SNPs were imputed by LinkImpute (Money et al., 2015). Command line scripts are supplied as Appendix.

Population structure

To visualize population structure of the tested fish, the *t*-SNE analysis was done following Chapter 1, section 1 with minor modification; software version of Python/scikit-learn was changed to 0.23.2. Command line scripts are supplied as Appendix.

Heritability and genetic correlation

Heritability of transformed HC and SL and genetic correlation between the two traits were estimated following Chapter 1, section 1. Command line scripts are supplied as Appendix.

Genome-wide association study (GWAS)

To search loci associated with transformed HC and SL, GWAS was performed following Chapter 1, section 1 with minor modification; the Bonferroni-corrected significant threshold was set as 5.40 ($= -\log_{10}(0.05/12{,}548)$). The effects of SNP markers were investigated by Bayes C model following Chapter 1, section 1. Command line scripts are supplied as Appendix.

Results

Artificial infection

The tested fish were produced from artificially crossing commercially raised fish (ten males and four females). The artificial infection trial was performed on those specimens at four months old for 46 days. Phenotypes were collected from 1,106 specimens. Average HC and SL were 15.85 (± 9.15 S.D.) and 9.83 (± 0.78 S.D.), respectively (Figure 2-1 and Supplementary Table S2-1). Figure 2-1 showed that the distribution of HC was non-normal (Shapiro–Wilk test: $p < 0.001$, alpha level = 0.05) while that of SL approximated a normal distribution (Shapiro–Wilk test: $p = 0.096$, alpha level = 0.05). Kendall's rank correlation analysis indicated no obvious phenotypic correlation between HC and SL ($\tau = 0.065$, $p = 0.002$). Since many zero values exist for HC, normal transformation was failed to adjust normality (i.e., square-root, logarithmic, and Box-Cox transformation were tested). Thus, raw data of HC and SL was used in the following genetic analysis.

Genotyping

An average of 4,033,985 (± 527,247 S.D.) raw reads per fish (n = 1,100) were generated by the Hiseq X sequencing. Then, the adapter sequences of raw reads were trimmed, and each fish had an average of 3,975,585 (± 460,483 S.D.) survived reads with average length of 133 bp (Supplementary Table S2-2). Subsequently, the survived reads were mapped onto a reference fugu genome (FUGU5/fr3). Following the SNP calling, filtration and imputation, 12,548 imputed SNPs were obtained for each fish.

Population structure

No clear clusters or strong stratification within the population was shown in Figure 2-2. The distribution of HC showed weak linear relationship between both x and y coordinate (Pearson's r = -0.112 and 0.117, respectively). This result ensured that biases from population structure was minimal or at least only limited in the following genetic analysis.

Heritability and genetic correlation

The small sample size and the low-density SNP panel of the pilot study in Chapter 1 might have reduced the credibility of genetic parameters (i.e., large standard error of heritability: HC: $h^2 = 0.308 \pm 0.123$ S.E.; SL: $h^2 = 0.405 \pm 0.131$). In this section, heritability for HC and SL were estimated using a medium number of samples with a medium density SNP set. As expected, the standard error of mean of the estimated heritability was narrowed down for each trait. However, the value for HC was decreased compared that estimated in the previous chapter ($h^2 = 0.162 \pm 0.041$ S.E.), while increased for SL ($h^2 = 0.537 \pm 0.045$ S.E.). The genetic correlation between HC and SL was also estimated using the same model as the previous chapter; similarly, a moderate undesired genetic correlation ($r_g = 0.245$) was observed between the traits (i.e., the larger the fish was, the larger the parasite loads was).

Genome-wide association study (GWAS)

To detect loci highly associated with HC and SL, GWAS was performed (Figure 2-3). For both traits, no SNP exceeded the significance threshold of 5.40 ($= -\log_{10}(0.05/12{,}548)$), suggesting these traits are polygenic as reported in Chapter 1. This is supported by a Bayesian approach (i.e., Bayes C) as each SNP has minimal effects (effect absolute value < 0.03) (Figure 2-4).

2.2 The effect of the number of SNPs on genomic prediction

The previous section confirmed the polygenetic nature of HC and SL, and thus the potential of GS that was suggested by the small-scale study in Chapter 1. In this section, I have estimated the predictive ability for the two traits using GBLUP, Bayes C and Bayesian RKHS. I have also examined the effects of the number of SNPs on the predictive ability of GBLUP model to determine the optimal SNP density for cost-efficient genotyping.

Materials and methods

Predictive abilities of genomic prediction

Predictive abilities of GP were obtained under three regression models, i.e., GBLUP, Bayes C, and Bayesian RKHS. The same models used in Chapter 1 were adopted. The cross-validation scheme was applied for calculation of predictive ability following Chapter 1, section 2 with minor modification; instead of tenfold cross-validation scheme for small-scale GP, fivefold cross-validation scheme was used in this analysis. Command line scripts are supplied as Appendix.

The effect of the number of SNPs on genomic prediction under GBLUP model

To examine the effects of SNPs size on GP for HC and SL, predictive ability was calculated by reducing the number of SNPs from 10,000 to 50 (10,000, 7500, 5,000, 2,500, 1,000, 500, 100, 50 SNPs) under GBLUP model. Specifically, a random sampling of tested SNP from the full SNP set (12,548) was performed using R/sample function with the default parameters for 50 times. In each time, fivefold cross-validation scheme was performed, and the predictive abilities were recorded. The average predictive abilities among 50 replicates were used to determine the optimal SNP density for GP. Details of cross-validation scheme were described in Chapter 1, section 1. To visualize the result, predictive abilities among various SNP density were plotted using R/ggplot2 package. Command line scripts are supplied as Appendix.

Results

The pilot study in Chapter 1, carrying out with a small number of samples and a low-density SNP panel, showed moderate predictive abilities of GP for both HC (0.307–0.325) and SL (0.460–0.463) under GBLUP, Bayes C, and Bayes RKHS. To test the generality of these results, the predictive abilities of GP were estimated using a population of a medium sample size and a medium-density SNP set under the three models. The predictive abilities (mean ± S.E.) were relatively low for HC (0.253 ± 0.008, 0.258 ± 0.007, and 0.256 ± 0.007, respectively) while modestly higher for SL (0.520 ± 0.007, 0.520 ± 0.007, and 0.524 ± 0.007, respectively) compared to those estimated in Chapter 1.

The result of GP for HC with various SNP sets (Figure 2-5 and Table 2-1) showed that average predictive abilities (± S.E.M.) of GP for HC and SL using 5,000 SNPs (HC: 0.256 ± 0.001; SL: 0.508 ± 0.008)

are comparable to those using 12,548 SNPs but decreased as the number of SNPs was less than 5,000. With 1,000 SNPs, the predictive abilities of GP for HC and SL showed 93.7% and 88.7% (HC: 0.242 ± 0.001; SL: 0.461 ± 0.008) of those estimated with the full SNP set.

Discussion

In this chapter, I used a middle-scale population of the tiger pufferfish and a medium-density SNP panel (i.e., 1,100 fish and 12,548 SNPs per individual) to examine the genetic architecture of heterobothriosis resistance (HC) and body size (SL). Then, I tested the feasibility of GS for these traits and determined the optimal SNP number for cost effective prediction using heterobothriosis resistance as the target trait. In all, these results supported the generality of the results obtained from the small-scale experiment in Chapter 1 (240 fish and 6,707 SNPs each individual).

With larger sample size and larger SNP density, moderate heritability was observed for SL ($h^2 = 0.537 \pm 0.045$) as in the case of the small-scale experiment ($h^2 = 0.405 \pm 0.131$); both results indicated the potential of GS for body size. On the other hand, the heritability for HC ($h^2 = 0.162 \pm 0.041$ S.E.) was relatively lower compared to that obtained in the small-scale experiment ($h^2 = 0.308 \pm 0.123$), presumably because the phenotypic distribution leans to zero in the middle-scale population. Nevertheless, the value still supports the possibility of GS for heterobothriosis resistance. Heritability estimation in this chapter has higher credibility (smaller S.E.) over Chapter 1 due to the larger sample size in the training population. Although phenotypic correlation between HC and SL was very small (Kendall's rank correlation analysis, $\tau = 0.065$, $p = 0.002$), their genetic correlation ($r_g = 0.245$) was moderate and similar as that estimated in the small-scale experiment ($r_g = 0.228$). Thus, as suggested in Chapter 1, breeders need to pay close attention to this undesired correlation especially when their breeding goal is to simultaneously improve both traits.

To confirm the availability of GS, GP was performed under the three models: GBLUP, Bayes C, and Bayesian RKHS. The predictive abilities were low for HC (0.253–0.258) while moderate for SL (0.520–0.524); modestly lower and higher compared to the those estimated in Chapter 1 (HC: 0.307–0.325; SL: 0.460–0.463), respectively; supporting the generality of the conclusion of the Chapter 1. Empirical experiment of GP in other aquaculture species showed the possibility of the use of low- or moderate-density SNP panel (few thousands to ten thousands SNP) for production traits (Hosoya et al., 2021; Kriaridou et al., 2020). As price of genotyping is generally associated with the number of SNPs, it is expected that the low-cost genotyping strategy can largely reduce the cost of GS and boost application of GS in aquaculture (Zenger et al., 2019). In this study, the result of GP for HC and SL with various SNP sets suggested 5,000 SNPs are sufficient for optimal GEBV estimation of these traits. Thus, genotyping cost of GS is highly expected to be reduced in the tiger pufferfish aquaculture.

Chapter 3. Availability of low fishmeal diets for the tiger pufferfish

In the previous chapters, it was confirmed that GS is available for improvements in heterobothriosis resistance and body size of the tiger pufferfish fed with a standard diet. In the following chapters, the feasibility of GS for these traits was investigated with LFM diets. Among various fishmeal replacements, such as plant-based ingredients, animal byproducts, and fishery byproducts, single cell proteins such as yeast and bacteria, recently though to be prospective protein sources which accomplish sustainable developments in aquaculture (Hua et al., 2019). As for the tiger pufferfish, while availability of soybean meal, meat bone meal, and freeze-dried blue mussel have been examined (Kikuchi and Furuta, 2009; Lim et al., 2011; Ukawa et al., 1996), it has not yet been examined if single cell proteins can replace fishmeal. In this chapter, I have tested the utility of LFM diets of which fishmeal contents were partially replaced with single cell proteins or plant protein, by comparing growth performance, physiological responses (blood chemical compositions and liver transcriptomes), and disease resistance against heterobothriosis between a regular fishmeal diet.

3.1 Growth performance

The most fundamental aspect of diet evaluation is to investigate growth performance of test fish since the growth performance is the most import production trait (Janssen et al., 2017) while fishmeal replacements may lack in important nutrients or even contain undesired substances that slow individual growth and developments. For example, plant protein typically have anti-nutritional elements, poor amino acids profile, and contain carbohydrates that are not digestible in carnivorous fish (Gatlin et al., 2007; Glencross et al., 2007; Hua et al., 2019). In particular, both yeast-based and bacteria-based ingredients contain high level of dietary nucleic acids, of which catabolites can cause hyperuricemia and negative effects on nitrogen metabolism, and thus growth performance as observed in rainbow trout (Perera et al., 1995; Rumsey et al., 1991) . In this section, I have conducted a three-month feeding trial to evaluate the effects of the following LFM diets on growth performances of the tiger pufferfish: three diets contained 37.5% replacement of fishmeal with plant-, bacteria-, or yeast-based ingredients; the fourth contained 62.5% replacement with yeast protein; a reference (control) diet with a high fishmeal content of 80% was used for comparison. The fish were fed one of these diets for three months and the effects on growth performance were then evaluated.

Materials and methods

Diets

All feeds used during the feeding trial were purchased from Hayashikane Sangyo Co. Ltd. (Shimonoseki, Yamaguchi Prefecture, Japan). The control (FM) diet had a fishmeal content of 80%. Three LFM diets contained either a plant (PP)-based, bacterial cell (BAC)-based, and yeast (LY)-based proteins, replacing 37.5% of fishmeal, i.e., containing 50% fishmeal and 30% replacement content, respectively. The fourth diet had a high yeast (HY)-based protein, consisting of 30% fishmeal and 50% yeast proteins. The chemical composition, energy, amino acid composition, and polyunsaturated fatty acids (PUFAs) of the five diets were analyzed by the Japan Functional Food Analysis and Research Center (Fukuoka, Fukuoka Prefecture, Japan) (Table 3-1).

Experimental fish production

Experimental fish were produced by a factorial crossing among 17 cultured females and 56 cultured males (898 crosses in total) and communally reared using a conventional culture method following Chapter 2. A total of 360 one-year-old individuals were sampled from a tank in which no fish attached by *Heterobothrium okamotoi* were observed (examined on ten individuals in the tank) and used for the following experiments. Until commencement of the feeding trial, the fish were fed with the standard expanded pellet (EP) diet that has a high fishmeal content (Lavelarva and Junior, Hayashikane Sangyo Co., Ltd., Yamaguchi Prefecture, Japan).

Feeding trial

The effects of the four LFM diets on growth performance were investigated in a feeding trial using 360 tiger pufferfish for three months (from February 14, 2019 to May 15, 2019). For the trial, 24 fish were placed in each of 15 circular one-ton experimental tanks. The experimental tanks were based in a 100-ton circular concrete tank filled with ambient sea water that acted as a water bath to keep the temperature the same in each tank. The tanks were aerated, provided with flow-through sea water, and provided with a regular photoperiod. The fish were fasted for one day prior to transfer from the communal tank to the experimental tanks. During transfer, the fish were anesthetized with 200 mg/l of 2-phenoxyethanol, and standard lengths (SLs) and body weight (BWs) were measured; the fish were also tagged with a passive integrated transponder (PIT) tag on the left-side of the body and were fin clipped for genetic analysis. Fin clips were stored in 800 μl of 100% EtOH at -30°C until use. The 15 tanks were divided into five groups of three, which were fed one of the five diets, i.e., FM, PP, BAC, LY, or HY. Fish were fed manually twice a day until apparent satiation. The amount of feed, water temperature, and dissolved oxygen were recorded for each tank. SLs and BWs were measured once per month; thus, including the initial measurements, four measurements were made during the trial (SL_n and BW_n for n = 1, 2, 3, 4). One fish died during the trial; this was in the FM group (died on April 5, 2019). Box plots of the SL and BW data were generated using R/ggplot2. All rearing records are available at Mendeley Data, doi: 10.17632/d94646df6p.1. To compare growth performance among dietary groups, growth rates relative to initial body size (%) for SL and BW (ΔSL_{41} and ΔBW_{41}) were calculated for each individual as follows:

$$\Delta SL_{41} = 100 \times (SL_4 - SL_1)/SL_1,$$

$$\Delta BW_{41} = 100 \times (BW_4 - BW_1)/BW_1$$

Statistical tests

The effects of low-fishmeal diets on relative growth rates were examined using generalized linear models (GLMs). The GLMs were fitted by the *glm* function implemented in R/stats (v4.1.0). I also tested if there were unexpected differences in initial body size (SL_1 and BW_1) between diet groups. For this analysis, model selection was carried out on the four models: a GLM with the diet groups as the fixed effect assuming Gaussian distribution, a GLM with the diet effect assuming Gamma distribution, a Gaussian GLM without the diet fixed effect, and a Gamma GLM without the diet effect. As for ΔSL_{41} and ΔBW_{41}, additional four models were constructed: a Gaussian or a Gamma GLM which included the tank effects as the fixed effect, and those with both diet effects and tank effects as the fixed effects (eight models in total).

To select the best fit model, the Akaike information criterion (AIC) was computed for each model using the *extractAIC* function in R/stats. In addition, ΔAIC, the difference between the minimum AIC and the tested AIC, and the Akaike weight (w) were computed using the *akaike.weights* function in R/qpcR (v1.4.1) (Ritz and Spiess, 2008). The GLM with the highest w was considered the best model. When the diet effects were supported by the model selection, a post hoc analysis was performed using R/emmeans (v1.6.6). In this analysis, the estimated marginal means (EMMs) for dietary treatment groups or experimental tanks were calculated using the e*mmeans* function, and a two-tailed pair-wise t-test (α = 0.05) was performed with p-value adjustment by the Tukey method for all comparisons using the *contrast* function with the augment *method = pairwise*.

Results

A three-month feeding trial was conducted on 360 cultured tiger pufferfish to evaluate the impacts of LFM diets on growth performance (Figure 3-1 and Table 3-2). Differences in SL_1, BW_1, ΔSL_{41}, and ΔBW_{41} among experimental groups fed and the control FM-fed group were tested using the GLM procedure (Table 3-3). The EMMs (with 95% confidence intervals) for SL and BW are shown in Figure 3-2. At the start of the feeding trial, although a Gaussian model with the diet term as the fixed effect was selected, no significant differences in either SL_1 or BW_1 were observed between experimental groups and the FM-fed group in the post hoc analysis (SL_1: p-value = 0.406–0.981; BW_1: p-value = 0.296–0.999). For the comparison of relative growth rates, GLMs with the diet effects assuming Gaussian and Gamma distribution were selected for ΔSL_{41} and ΔBW_{41}, respectively. The tests indicated that mean ΔSL_{41} and ΔBW_{41} in the HY-fed group (ΔSL_{41}: 20.68 ± 4.21 cm; ΔBW_{41}: 73.88 ± 19.06 g) were significantly greater than the FM-fed group (ΔSL_{41}: 18.71 ± 4.21; ΔSL_{41}: 63.21 ± 19.40) (p = 0.031 and 0.008, respectively).

3.2 Blood biochemistry

Blood biochemistry has been widely used as an indicator of health status of individual fish (Esmaeili, 2021). In this section, I have evaluated changes in the following blood biochemistry as well as the hematocrit level (Ht) of the tiger pufferfish right after the three-month feeding trial: the levels of total cholesterol (TCHO), triglyceride (TG), total protein (TP), and glucose (GLU) in plasma.

Materials and methods

Blood test

To examine the effects of LFM diets on blood biochemistry, blood samples were extracted from 150 randomly selected individuals (ten fish per experimental tank/ 30 fish per diet) immediately after the final measurement of body size. For each fish, approximately 1 ml of blood was taken from the caudal vein using a syringe with a 19G needle; samples were kept in 1.5 ml tubes on ice. Syringes, needles, and tubes were heparinized with 5,000 IU/μl heparin sodium. Subsequently, the liver of each fish was dissected to measure liver weight (LW) and calculate the hepatosomatic index (HSI = LW/BW_4*100). Gonadal sex was visually examined in each fish. Hematocrit tests were performed twice per fish: an aliquot of whole blood was drawn into a heparinized glass tube and sealed with clay. The blood-filled tubes were then placed in an H-1200F centrifuge (Kokusan, Tokyo, Japan) and spun at 926 × g for 15 min. After fractionation of the blood, the hematocrit level (Ht) was measured using a hematocrit scale. Blood plasma was separated from whole blood in a 1.5 ml heparinized tube by centrifugation at 926 × g for 15 min and stored at -80°C until analysis. The levels of total cholesterol (TCHO), triglycerides (TG), total protein (TP), and glucose (GLU) were analyzed in the plasma samples using a DRI-CHEM 7000V analyzer (Fujifilm, Tokyo, Japan).

Statistical tests

The effects of low-fishmeal diets on blood chemistry were tested using GLMs as described above. Model selection was performed on the eight models, models with and without the diet groups and the tank as fixed effects (i.e., only diet, only tank, diet and tank, or no fixed effect) assuming either Gaussian or Gamma distribution. Model selection and post hoc tests were performed as described above.

Results

Physiological effects of the LFM feeds were investigated from Ht and the blood biochemical composition (Figure 3-3 and Supplementary Table 3-4). The GLM procedure was applied for comparison (Table 3-5) and the EMMs with 95% CIs of those phenotypes were visualized in Figure 3-4. As for Ht, the Gaussian model including dietary treatment as the fixed effect was selected, and significant effects on Ht were observed in HY-fed group where this group showed lower average Ht (25.2 ± 5.1%) compared to the control group (29.3 ± 5.6%) (p = 0.009). In the similar manner, the Gaussian model with the fixed effect of dietary treatment was

chosen, and significant effects of diet group was observed in the average TCHO, where that of the BAC- (175.9 ± 48.1 mg/dl), LY- (173.8 ± 36.0 mg/dl), and HY-fed (179.6 ± 25.8 mg/dl) groups were lower than that of the control group (206.8 ± 44.1 mg/dl) (p = 0.015, 0.007 and 0.047, respectively). As for TG, GLU, TP and LW, while model selection suggested the effect of dietary group, no significant differences were supported by post hoc analysis. Regarding HSI, the model assuming Gamma distribution including dietary treatment as the fixed effect was chosen, and average HSI in BAC-fed group (12.1 ± 5.2 %) was significantly higher than that of the FM-fed group (10.3 ± 1.5 %) (p = 0.002).

3.3 Response to parasite infection

The fish fed with HY diets showed better growth rate compared to those fed with FM diet and the other LFM diets. Low fishmeal diets, however, may affect immune responses and reduce disease resistance. For instance, yellowtail *Seriola quinqueradiata* fed with a non-fishmeal diet (replaced with plant proteins) showed higher mortality caused by natural infection with pseudotuberculosis and artificial infection with *Lactococcus garvieae* (Maita et al., 2006). Moreover, red sea bream (*Pagrus major*) fed with a LFM diet (50% of fishmeal were replaced by soy protein concentrate) showed the lower resistance against bacterial infection caused by *Edwardsiella tarda* (Khosravi et al., 2015). As for the tiger pufferfish, individuals fed with HY diets showed reduces in Ht level, and this may also hinder the resistance against heterobothriosis as it is a hematophagous parasites (Ogawa, 2016). In this section, I have conducted an artificial infection trial right after the three-month feeding trial to compare the heterobothriosis resistance of the tested fish fed with normal diet and the four LFM diets.

Materials and methods

Artificial infection

To investigate the effects of the LFM diets on parasitic resistance against *H. okamotoi*, fish were artificially exposed to the oncomiracidium of the parasite following Chapter 1 with minor modifications. Fish were transferred into three circular experimental tanks of two metric ton capacity newly setup for a three-hour exposure to the oncomiracidium. To avoid sampling errors, ten fish from each dietary group were mixed in each experimental tank (50 individuals per tank) and kept overnight without feeding. Prior to the infection treatment, water depth was adjusted to 15 cm, and then approximately 5,000 oncomiracidium (100 oncomiracidium per fish) was introduced into each tank. At three-hour post exposure, fish were returned to their original tanks according to their PIT tag ID and reared using the same experimental diets as before until the measurements of the number of infected parasites. At 36-days post infection, SL and BW of each fish were measured (SL_5 and BW_5) and the branchial cavity walls (BCW) were dissected out from both sides following euthanasia with overdose 2-phenoxy ethanol (> 600 ppm). Meanwhile 10 fish obtained from one of the replicate tanks of each diet group were used to evaluate Ht as described above. The dissected BCW was kept in 10% formalin until counting the number of parasites under the microscope. The host resistance against *H. okamotoi* was accessed by the counts of adult parasites on the BCW (HC), and HC standardized by square of SL_5 (HD = 100 × HC/SL_5^2). To evaluate the growth performance for each dietary treatment group during the artificial infection, the growth rate in SL and BW relative to SL_4 and BW_4, respectively, (ΔSL_{54} and ΔBW_{54}, %) was calculated.

Statistical tests for the effects of diets on parasite loads

The effects of low-fishmeal diets on HC, HD, Ht, ΔSL_{54} and ΔBW_{54} were tested using GLMs as described above. Regarding HC, model selection was done on the eight models, models with and without the diet groups and the infection tanks as fixed effects (i.e., only diet, only tank, diet and tank, or no fixed effect) assuming either Poisson or negative binomial distribution. The GLM assuming a negative binomial distribution

was realized by *R/glm.nb* function. As for HD, model selection was done on the four models: Gaussian GLMs with and without the diet groups and the infection tanks fixed effects. I did not consider models under Gamma distribution for this trait since phenotypic values of some individuals were zero. Model selection for Ht was done on the eight models, models with and without the diet groups and the tank as fixed effects (i.e., only diet, only tank, diet and tank, or no fixed effect) assuming either Gaussian or Gamma distribution. Since one fish showed zero value of ΔSL_{54} and nine fish recorded non-positive values of ΔBW_{54} ($-8.0 \leqslant \Delta BW_{54} \leqslant 0$), only Gaussian GLMs with and without the diet group and the feeding tanks as the fixed effect were fitted. Model selection and post-hot test was done as described above.

Results

The effect of LFM diets on resistance to the parasite *H. okamotoi* was investigated using 150 fish in three replicate tanks (Figure 3-5 and Table 3-6). For the GLM procedure, the selection metrics (i.e., AIC, ΔAIC, and *w*) of GLM candidates are shown in Table 3-7. For HC and HD, models assuming negative binomial and Gaussian distributions, respectively, with only infection tank effects gave best fits, suggesting that there were no significant differences among the different diet groups. With regard to Ht, a model assuming a Gaussian distribution without a fixed effect was selected, indicating that impact of the different diets on Ht was negligible. The best model for a Gaussian distribution with dietary effect as the fixed effect for ΔSL_{54} indicated that the mean ΔSL_{54} in the LY-fed group (10.0 ± 2.6%) was significantly higher than in the FM group (6.9 ± 3.5%) (p = 0.001). By contrast, the Gaussian model with both dietary effect and feeding tank effect was selected for ΔBW_{54}, and the post hoc analysis indicated there was no significant difference between the diet groups and the control group ($p > 0.420$).

3.4 Transcriptomic responses in the liver

The knowledge of transcriptomic changes sheds light on the underlying molecular mechanisms of physiological adjustments by the fish to a LFM diet (Martin et al., 2016). Recently, transcriptomic profiling is available for aquaculture species to survey the genes or gene sets differentially expressed under the LFM treatment (Geay et al., 2011a; Kortner et al., 2013; Lazzarotto et al., 2018; Maita et al., 2006), and these information have been investigated for upgrading LFM diets using plant-, animal-, and insect-based proteins (Caballero-Solares et al., 2018; Geay et al., 2011; Kemski et al., 2020; Leduc et al., 2018; Martin et al., 2016). On the other hand, the use of novel diets based on single cell proteins is at an initial stage and only a few transcriptomic analyses have been carried out because of the costs of the analyses and the lack of a reliable annotated reference genome. For the tiger pufferfish, a high-quality annotated reference genome (Kai et al., 2011) is available as used in the previous chapters. In addition, short read RNA-sequencing which harnesses next-generation sequencing (NGS) technology largely reduces the cost and offers comprehensive, genome-wide transcriptomic information (Stark et al., 2019). Once the transcriptomes from different treatment groups of interested were sequenced, differential gene expression analysis could exploit the derived information and reveal the differences in abundance of gene transcripts (McDermaid et al., 2019). Subsequently, gene set analysis (e.g., gene set enrichment analysis (GSEA) and over representation analysis (ORA)) utilizes knowledge of biological pathways (Mathur et al., 2018) to further identify the gene sets with differential expression levels (Maleki et al., 2020; Mathur et al., 2018). These gene or gene sets with differential expression levels may be the fundamental genetic factors that interpret physiological responses to LFM treatment. In this section, I have performed an RNA-seq, differential gene expression analysis and gene set analysis using liver tissues to obtain a comprehensive insight into the physiological responses of this species to LFM diets.

Methods and materials

RNA sequencing

To assess the comprehensive physiological responses to each experimental diet, transcriptome in liver tissue was analyzed based on a 3′ mRNA Sequencing technology, RF-seq (Veeranagouda et al., 2019) with the modification, which adopted the multiplexing methods from BRB-seq (Alpern et al., 2019). This sequencing technology allows parallel sequencing of many samples at low cost because it is focused only on the 3′ ends of transcripts, and thus required fewer short reads per sample (Stark et al., 2019). Specifically, immediately after the blood extraction, three portions of liver tissue (approximately 100 mg each) were dissected and minced to store in 1.5 mL RNAlater stabilization reagent (QIAGEN, Hilden, Germany) at -20 °C. The total RNA was extracted using FastGene™ RNA Premium Kit (Nippon Genetics, Tokyo, Japan) with the DNase-treatment following the standard protocol of the manufacture's instruction. In the protocol of the modified RF-seq, each mixture containing 10 ng total RNA, 1 µL of 10 µM oligo(dT) attached with 10-bp unique molecular identifiers (UMI), 6-bp sample barcodes and P7 Illumina adapter sequences, and 4 µL of 5X First-Strand Buffer supplemented in SuperScript II Reverse Transcriptase (ThermoFisher, Kansas, USA) were incubated at 94 °C for 2.5 min for RNA fragmentation. The fragmented RNA was mixed with 2 µL of 10 mM dNTPs, 1 µL of 0.1 M DTT, 4 µL of 5 M Betaine, 0.12 µL of 1 M MgCl2, 1 µL of TSO (/5Me-isodC//iisodG//iMe-isodC/ GTC

TCG TGG GCT CGG AGA TGT GTA TAA GAG ACA rGrG+G), 0.5 µL of RNaseOUT (ThermoFisher), 1µL of SuperScript II Reverse Transcriptase (ThermoFisher), and 0.38 µL of RNase-free water, and reverse transcribed at 42 °C for 90 min followed by completion of reverse transcription consist of 10 cycles of 50 °C for 2 min and 42 °C for 2 min and inactivation of the reverse transcriptase at 70 °C for 15 min. The barcoded cDNAs were pooled and double-purified with 0.9X AMPure XP (Beckman Coulter, Brea, USA) and eluted in a 30 µL of RNase-free water. The elute was amplified in a mixture containing 12.5 µL 2xKAPA HiFi HotStart ReadyMix (Kapa Biosystems, Wilmington MA, USA), 1.25 µL each of 10 µM i7 index primer and P5 universal primer (Alpern et al., 2019), and 10 µL of the elute by following PCR program; initial denaturation at 98 °C for 3 min, 12 cycles of 98 °C for 10 sec, 63 °C for 30 sec and 72 °C for 30 sec, and final extension at 72 °C for 5 min. The sequence-ready library was obtained by purifying amplicon with 0.9X AMPure XP (Beckman Coulter) and quantified by qPCR using GenNext NGS Library Quantification Kit (TOYOBO, Osaka, Japan). Sequencing was done on Illumina NextSeq550 platform with paired-end mode using NextSeq v2 sequence kit (75 cycles) (Illumina, San Diego, USA). The forward reads (16-bp) consisting of the sequences sample barcodes and unique molecular identifiers (UMIs) while the reverse reads contained 68-bp of cDNA fragments. Each library was identified based on 8-bp Illumina i7 index sequence. Basecall was done using bcl2fastq on the NextSeq550 platform with *--minimum-trimmed-read-length 5* option. Three libraries (Lib1: n = 19; Lib2: n = 52; Lib3: n = 44) were generated (doi: 10.17632/d94646df6p.1); Lib1 was sequenced at Eurofin genomics (Tokyo, Japan) and the others were sequenced at National Institute for Basic Biology (Aichi, Japan).

NGS analysis

The sequence data were processed following the BRB-seq pipeline (Alpern et al., 2019). The forward sequences were trimmed by BRBseqTools (v1.6, Alpern et al., 2019) with augment: *TRIM*. The reference genome file and reference transcriptome file (fr3.fa and fr3.anotation.gtf) of FUGU5 (Kai et al., 2011) were downloaded using genomepy (v0.9.1) to generate STAR genome index by STAR (v020201) (Sibert and Nielsen, 2001) with the augments: *—runMode genomeGenerate*. Subsequently, the trimmed FASTQ file containing the reads of cDNA fragments was mapped onto the generated STAR genome index using STAR with the augments: *--runMode alignReads --outFilterMultimapNmax 1 --readFilesCommand zcat --outSAMtype BAM Unsorted*. The derived BAM files and the FASTQ files containing both barcodes and UMIs were used for the sample demultiplexing and the generation of transcriptome count/UMI matrix by BRB-seqTools with the following parameters: *CreateDGEMatrix -p BU -UMI 10*.

Data preprocessing and differential gene expression analysis

Data preprocessing and differential gene expression analysis was performed following Smyth et al (Smyth et al., 2018). R/edgeR (v3.34.0) (Chen et al., 2016) was used for the data preprocessing. At first, the transcriptome count matrix obtained as above was loaded using *DGElist* function. One sample with extremely low read counts was excluded and the genes that are lowly expressed was filtered out by *filterByExpr* function. Data normalization based on M-values (Robinson and Oshlack, 2010) was applied on the filtered data using

calcNormFactors function. Subsequently, the normalized data was transformed to $\log_2$(count per million) using *cpm* function with the augment: *log=TRUE*. To visually examine the biases among libraries, principal component analysis (PCA) plot and *t*-distributed stochastic neighbor embedding (*t*-SNE) plot were generated using *sklearn.decomposition.PCA* and *sklearn.manifold.TSNE* class of Python/scikit-learn (v0.23.2), respectively. The PCA was done with the default parameters. For *t*-SNE analysis, the perplexity was set as 20 and default values were used for the other parameters. As library bias was minimal (Supplementary Figure 3-6), the results of transcriptome count/UMI from three libraries were merged per sample for the following differential gene expression analysis.

To perform differential gene expression analysis, R/limma (v3.48.0) (Ritchie et al., 2015) was deployed. The *voom* function (Law et al., 2014) was applied to transform the normalized count to the $\log_2$(count per million), and to adjust the heteroscedasticity. Using *lmFit* function (Phipson et al., 2016), the gene-wise linear models (Smyth, 2004) were fitted. To compute the estimated coefficients and standard errors between low fish diet groups and the reference group, *contrasts.fit* function in R/limma (v3.48.0) with the augment *"contrasts"* to setup a custom contrast matrix. The matrix containing four comparisons, i.e., PP-FM, BAC-FM, LY-FM, HY-FM, was generated by *makeContrasts* function. To find the differentially expressed genes from each contrast, gene-wide empirical Bayes moderated *t*-tests (Smyth, 2004) were performed using *eBayes* function with the default augments. The results were visualized by Venn diagram using R/ggvenn (v0.1.9). As the stricter threshold, the log2 fold change threshold was applied using *treat* with the augment *"lfc = 1.2"* (McCarthy and Smyth, 2009). The gene symbols form HUGO gene nomenclature committee (HGNC) and zebrafish information network (ZFIN) were surveyed using R/biomaRt (v2.48.0, Ensembl). First, the ensemble dataset (v98) was connected using *useEnsembl* function and then the data set of FUGU5 was selected using *useDataset* function with the augment "*trubripes_gene_ensembl*". From FUGU5, the gene symbols were obtained using *getBM* function with the "*ensembl_gene_id*" and "*zfin_id_symbol*" attributes.

Gene set analysis

To identify the gene sets with differential expression levels, gene set enrichment analysis (GSEA) and over representation analysis (ORA) for Kyoto encyclopedia of genes and genomes (KEGG) pathways (Kanehisa et al., 2009) were performed using R/clusterProfiler (v4.0.0) (Wu et al., 2021; Yu et al., 2012). The KEGG pathways of *T. rubripes* were applied for KEGG-GSEA and KEGG-ORA. All genes were used for KEGG-GSEA. KEGG-GESA was conducted using *gseKEGG* function with the augments "*eps = 0, scoreType = pos*". The KEGG-ORA was performed based on the differentially expressed genes without the log2 fold change threshold using *enrichKEGG* functions. Results were visualized using R/ggplot2.

Results

NGS analysis

RNA-seq and differential gene expression analysis were performed to investigate physiological responses to the different diets. After processing of the BRB-seq pipeline, the derived transcriptome count

matrix included total 251,207,724 read counts from 115 individuals and 21,107 detected genes. The sequence reads have been deposited in the DDBJ Sequence Read Archive (DRA Accession: DRA012958).

Data preprocessing and differential gene expression analysis

Further filtration was performed to remove low-quality samples and transcriptome genes in the raw transcriptome count matrix, resulting in a read count of 248,935,561 for the 114 individuals and 7,950 detected genes. The filtered data was normalized and transformed to $\log_2$(count per million) values for the following analysis.

After fitting the gene-wise linear model with the preprocessed data, the results of empirical Bayes moderated *t*-statistics suggested that one gene was up-regulated in the PP-fed group compared to the control FM-fed group; 55 up- and 527 down-regulated genes were found in the BAC-fed group; eight genes were up-regulated in the LY-fed group; 13 up-regulated genes and four down-regulated genes were found in the HY-fed group. Among these differentially expressed genes, the gene *hmgcra* was up-regulated in all four LFM diets compared to the control (FM) group; an additional seven genes were up-regulated in both LY- and HY-fed groups (Figure 3-7). With the strict (log2 fold change) threshold, only the HY-fed group showed significant upregulation of *hmgcra*.

Gene set analysis

The gene sets with differential expression levels were explored in an enrichment analysis using KEGG-GSEA and KEGG-ORA for four comparisons, namely PP-FM, BAC-FM, LY-FM, and HY-FM (Figure 3-8, Figure 3-9, and Table 3-8). The KEGG-GSEA analysis revealed that "*Terpenoid backbone biosynthesis*" was over-represented in the LY-FM and HY-FM comparisons (adjusted *p*-value = 0.033 and 0.020), and "*steroid biosynthesis*" was significantly enriched in the PP-FM, LY-FM, and HY-FM comparisons (adjusted *p*-value < 0.001). The KEGG-ORA analysis also indicated that "*Terpenoid backbone biosynthesis*" was over-represented in the PP-FM, LY-FM, and HY-FM comparisons (adjusted *p*-value $\leq$ 0.004), while "*steroid biosynthesis*" was significantly enriched in the LY-FM and HY-FM comparisons (adjusted *p*-value < 0.001). Twenty gene sets were over-represented in the BAC-FM comparison; these included genes for amino acid biosynthesis/metabolism and immune responses.

Discussion

In this Chapter, I have evaluated the utility of four LFM diets in which fishmeal contents was partially replaced by proteins from other sources. The effects of these diets were examined with regard to growth performance, blood chemistry, resistance against the parasite *H. okamotoi*, and changes to the transcriptome in the liver. Overall, I found that a LFM supplemented with a high level of yeast protein (the HY diet) had the greatest potential for replacing the standard commercial FM diet, as the tiger pufferfish fed with the HY diet grew better than those fed with the FM diet and showed no severe deleterious effects.

At the end of the three-month feeding trial, HY-fed fish showed a significantly higher growth rate as adduced by SL ($p = 0.031$) and BW ($p = 0.008$) than FM-fed fish. Moreover, fish fed with the low yeast meal diet (LY) showed similar growth in SL and BW to FM-fed fish. The potential of yeast protein to be used as a replacement for fishmeal has been reported for several other fish species, namely, Atlantic salmon (Øverland et al., 2013), red sea bream (Takii et al., 2004), goldfish (*Carassius auratus*) (Gumus et al., 2016), rainbow trout (Hauptman et al., 2014; Vidakovic et al., 2020), and Nile tilapia (*Oreochromis niloticus*) (Ozório et al., 2012). However, a reduced growth rate was reported for some species after feeding with a high yeast diet, namely, rainbow trout (Hauptman et al., 2014; Vidakovic et al., 2020), Japanese flounder (*Paralichthys olivaceus*) (Kikuchi et al., 1993), channel catfish (*Ictalurus punctatus*) (Peterson et al., 2012), and Nile tilapia (Ozório et al., 2012). These results suggest that optimization of the level of yeast protein may be essential for maximizing the benefits of a yeast-based diet. The present study indicates that the tiger pufferfish tolerate the inclusion of yeast protein in the diet and that even 50% yeast protein enhances their growth performance.

Although the BAC-fed group showed a similar growth rate as the control group, they had a significantly larger hepatosomatic index (HSI). Several previous studies have reported successful use of diets containing low or medium levels of bacterial proteins in some species, namely, Atlantic salmon (Aas et al., 2006; Romarheim et al., 2011), rainbow trout (Perera et al., 1995), yellowtail (Biswas et al., 2020), and Florida Pompano (*Trachinotus carolinus*) (Rhodes et al., 2015). In these species, the growth performance was either maintained or enhanced compared to the standard diet. In the present study, the KEGG-GSEA analysis revealed that genes related to immunological responses were activated in the BAC-fed group; genes related to "*Salmonella infection*" were enriched only in this group (adjusted $p = 0.002$), even though *Salmonella* are not generally regarded as fish pathogens (Eng et al., 2015). In addition, the "*Endocytosis*", "*Cytokine-cytokine receptor interaction*", "*Phagosome*", and "*Tight junction*" categories (adjusted $p = 0.002$, 0.029, 0.030, and 0.029, respectively) were enriched; these pathways are associated with the invasion mechanism of *Salmonella* (Haraga et al., 2008; Jepson et al., 1995). These results argue against the long-term use of bacterial proteins as a protein source for the tiger pufferfish.

At the end of the feeding trial, HY-fed fish showed a decrease in mean Ht (25.2 ± 5.1%) compared to the control group (29.3 ± 5.6%, $p = 0.009$), suggesting the occurrence of mild hemolytic anemia. Similarly, hemolytic anemia was observed in rainbow trout fed with a diet in which 60% of the fishmeal protein was replaced by yeast protein (Huyben et al., 2017). In general, yeast contains a high level of nucleic acids, mainly RNAs, and catabolism of these dietary nucleic acids may cause a high concentration of plasma uric acids in fish; these might have an effect on protein, fat, carbohydrate, and uracil metabolism (Rumsey et al., 1991). Although

fish can catabolize plasma uric acids, the process can generate hydrogen peroxide, a reactive oxygen species that may damage erythrocytes and cell membranes if antioxidant levels are insufficient (Huyben et al., 2017; Sánchez-Muniz et al., 1982). In the liver transcriptomic analysis (KEGG-GSEA), I found that expression of genes related to "*Peroxisome*", "*Metabolism of xenobiotics by cytochrome P450*", "*Pentose and glucuronate interconversions*", and "*Purine metabolism*" were upregulated in the HY-fed group compared to the control group (adjusted p = 0.008, 0.031, 0.031, and 0.033, respectively). Peroxisomes are membrane-bound organelles that carry out oxidative reactions and contain a range of enzymes (Cooper, 2000). In the marine fish liver, peroxisomes contain enzymes that degrade uric acids, i.e., uricase, allantoinase and allantoicase (Hayashi et al., 1989; Noguchi et al., 1979). In the BAC-fed fish, the over-represented gene set "*Metabolism of xenobiotics by cytochrome P450*" includes four glutathione-related genes, *gstt1a* (glutathione S-transferase theta 1a), *gstt2* (glutathione S-transferase theta 2), *LOC101074398* (glutathione S-transferase Mu 3-like), and *LOC101065820* (glutathione S-transferase A-like). Using glutathione as a substrate, hydrogen peroxide can be reduced to water by catalase or glutathione peroxidase (Buetler et al., 2004). This gene set also contains three genes related to uridine 5'-diphospho (UDP)-glucuronosyltransferases, i.e., *ugt1a1* (UDP glucuronosyltransferase family 1 member A1), *LOC101076061* (UDP-glucuronosyltransferase 2C1-like), and *LOC101078725* (UDP-glucuronosyltransferase 1-1-like). Notably, *ugt1a1* produces the bilirubin uridine diphosphate glucuronosyl transferase enzyme that can glucuronidate bilirubin, a substance produced when erythrocytes break down (Fujiwara et al., 2016). The "*pentose and glucuronate interconversions*" pathway is able to regulate UDP-glucuronosyltransferase 1A1 (Sun et al., 2018). Therefore, it is hypothesized that the high level of dietary nucleic acids in a yeast-based diet can cause oxidative stress in the fish liver and also damage erythrocytes in blood plasma. However, Ht levels in the HY-fed group were stable even after a month of artificial heterobothriosis infection, and thus the effect does not appear to be unacceptably severe.

The mean TCHO level in the blood plasma of the HY-fed group (179.6 ± 25.8 mg/dl, mean ± S.D) was lower than that of the FM-fed group (206.8 ± 44.1 mg/dl). Cholesterols are essential lipid components of membranes and are vital for maintaining physiological functions in fish. Fish obtain cholesterols from both food intake and *de novo* synthesis (Norambuena et al., 2013). Natural yeasts do not synthesize cholesterol (Kitson et al., 2011); thus fish fed a reduced fishmeal diet that includes replacement yeast protein may have a reduced food intake of cholesterol. The *de novo* synthesis of cholesterol may possibly be enhanced to compensate for this insufficiency. This transcriptomic analysis of liver showed up-regulation of *hmgcra* in the HY-fed group compared to the FM-fed group. This gene regulates activity of hydroxymethylglutaryl-CoA reductase (Thorpe et al., 2004), a rate-limiting enzyme in the mevalonate pathway that has an important role in cholesterol and terpenoid/isoprenoid synthesis (Friesen and Rodwell, 2004); up-regulation of *hmgcra* may enhance cholesterol synthesis. The analyses using the relaxed threshold (i.e., without the log2 fold change criterion) identified *hmgcra* upregulation in PP-, BAC- and LY-fed groups; the mean plasma TCHO levels of these groups were less than that of the FM-fed group (and not significantly so in the PP-FM comparison). Previous studies on transcriptomes using qPCR and microarray profiling of fish fed no fishmeal or a LFM diet with plant-based replacement also identified *hmgcra* up-regulation: Atlantic salmon (*Salmo salar L.*) (Kortner et al., 2013), rainbow trout (*Oncorhynchus mykiss*) (Lazzarotto et al., 2018), yellowtail (*Seriola quinqueradiata*) (Maita et al., 2006), European sea bass (*Dicentrarchus labrax*) (Geay et al., 2011), and yellow perch (*Perca flavescens*) (Kemski et al., 2020). In addition, the gene set analysis in this study supported the importance of *hmgcra* and

other genes with "*Terpenoid backbone biosynthesis*" and "*steroid biosynthesis*" activities to physiologically adjust the cholesterol level under the LFM diet conditions. Adding dietary cholesterol supplementation to soya bean meal-based diets increased plasma TCHO values in rainbow trout; the supplementation increased cholesterol obtained by food intake and reduced the *de novo* synthesis of cholesterol, which is expensive in terms of energy (Deng et al., 2013; Norambuena et al., 2013). Likewise, adding dietary cholesterol supplementation in the HY diet might improve the plasma cholesterol level and further enhance growth performance of the tiger pufferfish.

Feeding fish with a LFM diet might impair their immune responses. I tested whether this was the case by artificially infecting the tiger pufferfish with heterobothriosis. No significant differences in parasitic loads were identified among the different diet groups. However, the distribution of HC and HD was relatively larger in the PP- and the HY-fed groups than the FM-fed group (Figure 3-5, a and b). This wider distribution suggests that the inter-individual differences in resistance to parasite infection may have been amplified by the PP and HY diets (i.e., gene-environment interaction). Heterobothriosis resistance in the tiger pufferfish has a genetic basis and the possibility of selective breeding for this trait has been suggested in Chapter 1 and previous studies (Kim et al., 2019; Lin et al., 2020). Theoretically, the larger the phenotypic variance, the larger the scope for genetic improvement by selective breeding. Therefore, my results suggest that the gains from selective breeding of the tiger pufferfish for increased heterobothriosis resistance will be greater in fish fed with PP or HY diet compared to the standard fishmeal diet.

In summary, the LFM diet HY containing 50% yeast protein has a high potential for replacing the commercial diet FM since the tiger pufferfish fed with HY did not show severe deleterious effects. The low Ht values of HY-fed fish are possibly caused by the high nucleic acids in HY, but there were no serious consequences regarding the health or growth of the fish. The increases in the expression levels of *hmgcra* and other genes related to "*Terpenoid backbone biosynthesis*" and "*steroid biosynthesis*" pathways suggesting compensatory enhancements in de novo cholesterol synthesis; it is anticipated that supplementation of cholesterol in the HY diet might improve the plasma cholesterol level and further enhance growth performance of the tiger pufferfish. This study showed that transcriptome information is a powerful approach for diagnosing and elucidating underlying molecular mechanisms to develop new diets for the tiger pufferfish. To further investigate the feasibility of GS for LFM tolerance of the tiger pufferfish, I attempted GP for heterobothriosis resistance and growth in SL using the LFM diet, HY, in the following chapters.

Chapter 4.: Genomic prediction for heterobothriosis resistance and body size under a short-term treatment of low fish meal diet

In Chapter 3, fish fed with the low fishmeal (LFM) diet of which 62.5% of fishmeal was replaced by yeast protein (HY) showed better growth compared to the regular fishmeal diet and the other LFM diets without severe deleterious effects. Moreover, larger phenotypic variation in *H. okamotoi* counts was observed among HY-diet fed fish compared to that among the FM-diet fed fish (HY: 25.2 ± 18.2; FM: 23.5 ± 12.5, mean ± S.D.); this might be caused by differences in inter-individual (genetic) responses to the different protein sources (i.e., gene-environment interactions), which possibly have impacts on heritability in heterobothriosis resistance. In other words, the genetic architecture and predictive abilities of GP for heterobothriosis resistance of the HY-fed fish may differ from those of the fish fed with the standard feed in Chapter 1 and Chapter 2; there might be a large effect QTL for the trait and MAS could be a better choice rather than GS. This can also be the case for growth traits. Therefore, I examined the genetic architecture and availability of GS for heterobothriosis and body size in the tiger pufferfish reared under a dietary treatment of HY diet. In this chapter, two populations were subjected to HY-diet treatment for three and four months, respectively, and artificial infection subsequently.

4.1 Genetic dissection of heterobothriosis resistance and body size traits

In this section, I performed a short-term feeding treatment and a subsequent artificial infection trial on two middle-scale populations (Pop-C and Pop-CW). After obtained phenotypes and genotypes, I calculated heritability for heterobothriosis resistance and body size under the dietary treatment. Then, I performed GWAS to examine the genetic architecture of both traits.

Materials and methods

Test populations and feeding treatments

All experiments including fish production, fish culture, the feeding treatment, the artificial infection trial, phenotyping, and sample collection were performed at Nagasaki Prefectural Institute of Fisheries (NPIF, Nagasaki, Japan). The artificial mating and fish culture follow Chapter 2 with minor modification. Two experimental populations were produced by means of artificial mating; Pop-C (n = 1,047) was generated by a total of 807 crosses among commercial broodfish (35 males and 29 females), while Pop-CW (n = 1,050) was produced from a total of 120 crosses among commercial broodfish (20 males and 19 females) and wild individuals (three males and five females) (Supplementary Table S4-1 and Supplementary table S4-2). Before starting the feeding treatment, fish were fed with standard EP diets that has a high fishmeal content (Lavelarva and Junior, Hayashikane Sangyo Co., Ltd., Yamaguchi Prefecture, Japan). A short-term dietary treatment with the HY diet was performed for 99 days (Pop-C) and for 128 days (Pop-CW) on approximately three-month-old fish. At the end of the feeding treatment, SL of total 993 and 1,017 fish were recorded for Pop-C and Pop-CW, respectively.

Artificial infection

Right after the short-term feeding treatment with HY diet, fish were subjected to an artificial infection following Chapter 1 with modifications: the set-up of experimental tanks for three-hour exposure to oncomiracidium of *H. okamotoi*, number of oncomiracidium introduced for each tank, and duration of fish culture post the exposure were changed. Specifically, for Pop-C, 15 identical one-ton experimental tanks were prepared (70 fish/tank), approximately 3,500 oncomiracidia were introduced into each tank (around 50 oncomiracidium per fish) for three-hour exposure, and the infected fish were reared for 36-37 days; for Pop-CW, 14 identical 0.5-ton experimental tanks (75 fish/tank), approximately 3,750 oncomiracidia per tank (around 50 oncomiracidium per fish) and reared for 41 days. Fish were fed with HY diet throughout the infection trial. At the end of the artificial infection trail, the branchial cavity walls of each fish were dissected and kept at -20 °C. These tissues were thawed for counting attached *H. okamotoi* (HC) under a stereo microscope.

Box-Cox transformation

To adjust the normality of phenotypes for phenotypic and genetic analysis, Box-Cox transformation (Box and Cox, 1964) was applied using R/caret(v6.0.90) (Kuhn, 2008) package. Specifically, Box-Cox transformation was calculated as follows:

$$y_i^{(\lambda)} = \begin{cases} \frac{y_i^{\lambda}-1}{\lambda}, & \lambda \neq 0 \\ logy_i, & \lambda = 0 \end{cases},$$

where $y_i^{(\lambda)}$ and y_i are transformed and raw phenotype for individual i, respectively; λ is an unknown parameter determined by *BoxCoxTrans* function with the default augments. Once the best λ was determined, $y_i^{(\lambda)}$ was calculated using *predict* function. Command line scripts are supplied as Appendix.

Genotyping

All procedures for genotyping, i.e., genomic DNA extraction, GRAS-Di sequencing, and the post genotype calling, were performed as described in Chapter 2 with minor modification; instead of HiSeq X Ten system, sequencing was done on a NovaSeq 6000 system using NovaSeq 6000 S4 reagent kit (Illumina, San Diego, USA). GRAS-Di sequencing was performed by Eurofin genomics Inc (Tokyo, Japan). Command line scripts are supplied as Appendix.

Population structure

Population structure of the tested fish was analyzed using the *t*-SNE analysis following Chapter 1 with minor modification, i.e., software version of Python/scikit-learn was changed to 0.23.2. Command line scripts are supplied as Appendix.

Heritability and genetic correlation

Heritability and genetic correlation were calculated using multivariate GBLUP model following Chapter 1. Command line scripts are supplied as Appendix.

Genome-wide association study (GWAS)

Genomic loci associated with transformed HC and SL under a short-term treatment of HY diet was searched by GWAS following Chapter 1 with a minor modification: the Bonferroni-corrected significant threshold was set as $\alpha = 2.60 \times 10^{-6}$ (0.05 divided by the number of SNPs: 19,227) for Pop-C and $\alpha = 2.35 \times 10^{-6}$ (divided by 21,272 SNPs) for Pop-CW. The effects of SNP markers were also investigated by Bayes C model following Chapter 1. Command line scripts are supplied as Appendix.

Results

Tested population and artificial infection

Two populations were subjected to the HY-diet treatment and the artificial infection trial. For fish from Pop-C (n = 993), average HC and SL were 12.52 (± 7.02 S.D.) and 17.64 cm (± 1.39 cm S.D.), respectively (Figure 4-1 and Supplementary Table S4-3), while the average HC and SL were 24.75 (± 14.64 S.D.) and 18.84 cm (± 1.66 cm, S.D.), respectively, for fish from Pop-CW (n = 1,017) (Supplementary Table S4-4). The distribution of HC was non-normal for both of the populations (Shapiro–Wilk test: $p < 0.001$ for both populations) (Figure 4-1). Therefore, HC was approximated to a normal distribution by means of Box-Cox transformation (HC+1 was used to avoid zero and λ was set to 0.6 for both populations). Moderate phenotypic correlation between HC and SL were observed (Pop-C: Pearson's $r = 0.489$, $p < 0.001$; Pop-CW: $r = 0.566$, $p < 0.001$). Transformed HC and SL was used in the following genetic analysis.

Genotyping

For Pop-C (n = 1,047), an average of 5,832,769 (± 561,757 S.D.) raw reads were generated by the NovaSeq 6000 sequencer. After trimming, an average of 3,579,364 (± 325,542 S.D.) reads were survived at both ends (Supplementary Table S4-5). For Pop-CW (n = 1,017), the average raw reads and survived reads were 5,857,528 (± 590,653 S.D.) and 3,670,410 (± 300,814 S.D.), respectively (Supplementary Table S4-6). The survived reads were mapped onto a reference fugu genome (FUGU5/fr3). Following the SNP calling, filtration and imputation, 19,227 and 21,272 imputed SNPs were obtained for each fish from Pop-C and Pop-CW, respectively. Fish with both genotypes and phenotypes (Pop-C: 993 fish; Pop-CW: 1,017 fish) were used for the following analysis.

Population structure

No strong stratification was observed for both of the populations (Figure 4-2). This result ensured that biases from population structure was minimal or at least only limited in the following genetic analysis. On the other hand, small clusters were observed in Pop-CW (Figure 4-2 (**b**)). In the t-SNE plot, individuals within the same cluster are genetically close and most presumably full-sibs. The parental fishes of Pop-C consisted of 64 farmed individuals sampled from approximately 10 farmed families, and thus included full- and half-sibs within the parental fish; siblings were genetically related at some extent (Supplementary Table S4-1). Contrary, the Pop-CW were produced from both commercial broodfish and a small number of wild individuals (three males and five females); a total of 47 parental fish came from 20 different farmed families as well as wild fish. Therefore, most of these parental fish were genetically unrelated, so were the siblings from different clusters (Supplementary Table S4-2).

Heritability and genetic correlation

The genetic parameter estimations for transformed HC and SL were performed on Pop-C and Pop-CW. For Pop-C, moderate heritability was observed for transformed HC (h^2 = 0.472 ± 0.051 S.E.) and SL (h^2 = 0.620 ± 0.046 S.E.), respectively. Similar results were obtained for Pop-CW: transformed HC, h^2 = 0.569 ± 0.052 S.E., and SL, h_2 = 0.669 ± 0.046 S.E.. With the same model, a moderate antagonistic genetic correlation (r_g = 0.661 and 0.585 for Pop-C and Pop-CW, respectively) was observed between the two traits.

Genome-wide association study (GWAS)

GWAS was applied to detect loci highly associated with transformed HC and SL in Pop-C and Pop-CW (Figure 4-3). One significant SNP (Chr3: 5,531,819, *ube4b*) was detected for HC in Pop-C. However, the effect of the locus on HC was only weak as HC of those individuals carrying two minor alleles at this SNP site (14.63 ± 7.02 S.D.) was scarcely higher than individuals carrying two reference alleles (11.30 ± 6.62 S.D.) as well as the population average (12.52 ± 7.02 S.D.). No significant SNP was found for SL in Pop-C or both of the traits in Pop-CW. These results (i.e., there were no large effect QTL for HC and SL of individuals fed with HY diet) were supported by a Bayesian regression approach (i.e., Bayes C) as each SNP had minimal effects (effect absolute value < 0.1) (Figure 4-4). These results also imply that both HC and SL of individuals experienced the short-term dietary treatment of HY feed are polygenetic as in the case of those fed with a standard feed.

4.2 Predictive abilities of genomic prediction

The feasibility of GS for heterobothriosis resistance and body size of the tiger pufferfish reared with a standard feed were revealed in Chapter 1 and 2. The moderate heritability and polygenetic nature detected in the previous section suggests GS is also helpful for the two traits of the individuals which reared with HY diet; this is tested in this section.

Materials and methods

Predictive abilities of genomic prediction

The phenotype and genotype data obtained in the previous section (Pop-C: n = 993; Pop-CW: n = 1,017) were used for estimation of predictive abilities of GP under three regression models: GBLUP, Bayes C, and Bayesian RKHS. Predictive ability was estimated via a fivefold cross-validation scheme as described in Chapter 2. Command line scripts are supplied as Appendix.

Results

The predictive abilities estimated under the three models, GBLUP, Bayes C and Bayesian RKHS, were similar and these values were high for both HC (Pop-C: 0.520–0.522; Pop-CW: 0.564–0.567) and SL (Pop-C: 0.601–0.604; Pop-CW: 0.736–0.740) (Table 4-1). These results indicated the feasibility of GS for the parasite resistance and the growth performance under a short-term treatment with HY diet.

Discussion

In this chapter, I have examined heritability and the genetic architecture for the heterobothriosis resistance and growth in SL of the tiger pufferfish subjected to a short-term feeding treatment with the low fishmeal diet of which 50% of biomass was consisting of yeast protein (HY). Then, I calculated the predictive abilities of GP for these traits under three different models. As the results, moderate heritability was obtained for both of the traits. Genome-wide association study revealed that these traits are polygenic. Moreover, the high predictive abilities indicated feasibility of GS for HC and SL of the species experienced the short-term dietary treatment of HY.

The estimated heritability (± S.E) for transformed HC and SL were moderate in both of the populations (Pop-C, transformed HC: $h^2 = 0.472 \pm 0.051$, SL: 0.620 ± 0.046; in Pop-CW, transformed HC: 0.569 ± 0.052, SL: 0.669 ± 0.046). The estimated heritability for each trait were even higher than those estimated in Chapter 1 (transformed HC: 0.308 ± 0.123; SL: 0.405 ± 0.131) and Chapter 2 (HC: 0.162 ± 0.041; SL: 0.537 ± 0.045), where tested individuals were fed with a standard feed. These differences are partly due to the differences in population structure of the experimental fish and/or the activities of the parasite. However, the results imply that contribution of genetic effects is larger for the fish fed with HY diet as the parental individuals of Pop-C and the population used in Chapter 2 were derived from the same commercially cultured population. Meanwhile, the predictive abilities estimated under the three models, i.e., GBLUP, Bayes C, and Bayesian RKHS, were also high for both HC (Pop-C: 0.520–0.522; Pop-CW: 0.564–0.567) and SL (Pop-C: 0.601–0.604; Pop-CW: 0.736–0.740); these values were higher than those estimated in Chapter 1 (transformed HC: 0.307–0.325; SL: 0.460–0.463) and Chapter 2 (HC: 0.253–0.258; SL: 0.520–0.524). Again, these results indicated the feasibility of GS for parasite resistance and growth performance under short-term HY diet feeding.

The gene-environment interaction have been widely studied in the field of breeding science, since production environments are often different from each other, while even small gene-environment interaction could affect breeding programs (M.E. Goddard and B.J. Hayes, 2007). Therefore, the regression models include the effect of gene-environment interactions can increase higher selection accuracy (Crossa, 2012; Crossa et al., 2017; Desta and Ortiz, 2014). In plant breeding, it is common that quantitative production traits, such as yields and plant weight, have different performance under different environmental factors, e.g., climate, soil, and managements (Iwata et al., 2016; Teressa et al., 2021; Xu et al., 2020). In case of livestock, this gene-environment interaction was also commonly observed (Mulder and Bijma, 2005; Rauw and Gomez-Raya, 2015). As for aquaculture species, the gene-environment interaction was documented in major species such as tilapia (*Oreochromis niloticus*), rainbow trout (Sae-Lim et al., 2015, 2013), and Pacific white shrimp (Castillo-Juárez et al., 2007), while low gene-environment interactions under the treatment of plant-based diets were realized in rainbow trout (Le Boucher et al., 2011; Pierce et al., 2008) and European sea bass (Le Boucher et al., 2013). Although the effect size of gene-environment interaction was not explicitly quantified in this study, the changes in estimated genetic parameters and predictive abilities of GP imply unignorable effects of gene-environment interactions. The results of this chapter showed that a breed of the tiger pufferfish which shows better growth and/or parasite resistance with HY-diet can be produced by means of GS. This would be possible for other species using a particular LFM diet as well. Furthermore, establishments of a robust breed which have resilience to multiple LFM diets would be theoretically possible with following steps: collect phenotypic data for each diet,

perform GP using all data adding gene-environment effects in prediction model, and select elite individuals based on GEBVs.

In summary, the high estimated heritability and predicted abilities of GP for these traits suggest that larger genetic gain can be expected for the individuals reared with HY diet compared to those fed with the standard diet. This possibility was further investigated by subjecting cultured population to a long-term feeding treatment in Chapter 5

Chapter 5. Genomic prediction for body size under a long-term treatment of low fish meal diets

In Chapter 4, I confirmed feasibility of GS for heterobothriosis resistance and body size of the tiger pufferfish fed with HY diet for a short period of several months. However, a long-term treatment of HY-diet may result in different outcomes. This short term treatment would be long enough to assess the availability of genetic improvement for heterobothriosis resistance because the most severe infection occurs at the early phase of farming, i.e., just after transfer from land-based hatcheries to sea cages. On the other hand, growth performance reflected as standard length (SL) should be evaluated at the harvest size (20–22 months for the species), and thus the feasibility of GS should be tested by a long-term treatment. However, evaluation at harvest is time-consuming; if GEBVs of SL at harvest can be predicted at an early phase, cost-efficiency of the breeding project can be dramatically improved, as it is often the case for selective breeding in highly prolific aquaculture species that the training population and selection candidates are full-/half-siblings produced at the same time.

The importance of genes related to terpenoid backbone biosynthesis and steroid biosynthesis for utilization of LFM diet was suggested in the gene expression analysis in Chapter 3. Thus, it is expected that predictive ability of GP can be improved by selecting the genomic variants around and within these differentially expressed genes from genome-wide SNPs, for example, the regression models with weighted genomic relationship matrix; large-effect genomic variants are up-weighted to construct this matrix, effectively improved the predictive ability of GP for production traits in US Holstein dairy cattle (Tiezzi and Maltecca, 2015). In this chapter, I have investigated the genetic basis and predictive ability of GP for SL under a long-term treatment of HY diet using 936 individuals which shared the same parental individuals with Pop-C from Chapter 4. I have also investigated the phenotypic and genetic correlation for SL at different time points using same individuals to test the possibility of prediction with a short rearing period. Finally, I have taken the advantage of the transcriptomic information for GP under a regression model with weighted genomic relationship matrix.

5.1 Genetic dissection of standard length under a long-term treatment of a low fishmeal diet

Materials and methods

Tested population and low fish meal diet treatment

The artificial mating and a long-term feeding treatment were performed in Nagasaki Prefectural Institute of Fisheries (NPIF, Nagasaki, Japan). The test population was derived from the same population as Pop-C in Chapter 4, crossing commercial broodfish (35 males and 29 females, 807 crosses in total) (Supplementary S4-1). Before the start of the feeding treatment, fish were fed with the standard EP diet that has a high fishmeal content (Lavelarva and Junior, Hayashikane Sangyo Co., Ltd., Yamaguchi Prefecture, Japan). Specimens (n = 936) were fed with HY for 19 months from juvenile (three-month old) to the harvest size, during which standard length (cm) was recorded four times for each fish: SL_1 (198th day of the long-term feeding treatment), SL_2 (319th day), SL_3 (422nd day), and SL_4 (569th day).

Genotyping

The genomic DNA extraction, GRAS-Di, and post analysis were performed as described in Chapter 4 with minor modification; instead of trimmomatic software, raw reads were trimmed by Trim Galore-0.6.6 (Krueger, n.d.), the wrapper script around Cutadapt-2.6 (Martin, 2011) and FastQC-0.11.9, with the parameters: *—length 100 –paired –stringency 10*. Fish in the test population (n = 1,000) and their parents (n = 84, including half-sibs) were genotyped together. Library preparation and sequencing were done by Eurofins Genomics Inc. (Tokyo, Japan). Command line scripts are supplied as Appendix.

Population structure

To visualize population structure within the test population, the *t*-SNE analysis was done following Chapter 1, but the software version of Python/scikit-learn was changed to 0.23.2 and the color of each circle represents SL_4. Command line scripts are supplied as Appendix.

Heritability and genetic correlation

Heritability and genetic correlation were calculated for SL_i following Chapter 1. Note that GP under multivariate model for longitudinal data (multiple measurements of a single trait across years) is regularly used for livestock (Strucken et al., 2015). Command line scripts are supplied as Appendix.

Genome-wide association study (GWAS)

To investigate the associated makers with SL_i, GWAS was performed following Chapter 1, where Bonferroni-corrected significance threshold was set as 5.518 (= – log10 (0.05/16,471)). Command line scripts are supplied as Appendix.

Results

Tested population and low fish meal diet treatment

Fish (n = 1,210) were subjected to a long-term feeding treatment with HY diet for 19 months. During the feeding treatment, standard length (cm) of a total of 936 individuals were recorded for four times (SL_i, i = 1,2,3, and 4). Average values of SL_1, SL_2, SL_3, and SL_4 were 24.18 (± 1.70 S.D.), 26.29 (± 1.84), 30.69 (± 2.13), and 35.56 (± 2.55) cm, respectively (Figure 5-1 and Supplementary Table S5-1). High Pearson's r between each pair of SL_i was observed (r = 0.849–0.947) (Table 5-1).

Genotyping

An average of 4,529,807 (± 494,169 S.D.) raw reads per fish (n = 1,084: 1,000 fish from the tested population and 84 from their parents) were generated by NovaSeq 6000 sequencing system. After trimming, each fish had an average of 3,557,154 (± 344,916 S.D.) survived reads (Supplementary Table S5-2). Subsequently, the survived reads were mapped onto a reference fugu genome (FUGU5/fr3). Following the SNP calling, filtration and imputation, 16,471 imputed SNPs were obtained for each fish. Since 936 individuals have full record of SL measurements, the derived genotypes of 936 fish were used for the following analysis.

Population structure

As seen in Figure 5-2, clear clusters or strong stratification within the tested population were not observed. This result ensured limited biases from the population structure to the results of heritability estimations, GWAS and GP.

Heritability and genetic correlation

Heritability for SL at harvest as well as three different time points of individuals fed with HY-diet were estimated. Moderate heritability was observed for each time point and the value gradually increased as the fish grow out (SL_1: 0.548 ± 0.050; SL_2: 0.609 ± 0.048; SL_3: 0.612 ± 0.048; SL_4: 0.613 ± 0.048, h^2 ± S.E.). Using the same model, a high genetic correlation (r_g = 0.917-0.973) was observed between each SL measurement (Table 5-2).

Genome-wide association study (GWAS)

GWAS was applied to detect loci highly associated with SL_i (for i = 1,2,3 4) (Figure 5-3). None of the loci exceeded the significance threshold.

5.2 Genomic prediction and correlation between GEBVs of body size at different time points

The previous section revealed moderate heritability for and polygenic nature of SL of the fish fed with HY-diet for 19 months at the for different time points including harvest (SL_1–SL_4). In this section, I have estimated the predictive ability for these SL measurements using GBLUP, Bayes C and Bayesian RKHS. Then, I further investigate correlation between GEBVs for the four measurements.

Materials and methods

Predictive abilities of genomic prediction

To study the effect of feeding HY-diet for a long time period (19 months), predictive abilities of GP for the four SL measurements were calculated under the three regression models, i.e., GBLUP, Bayes C and Bayesian RKHS. The cross-validation scheme was applied for calculation of predictive ability following Chapter 2. Command line scripts are supplied as Appendix.

Correlation between GEBVs for SL at different time points

To examine correlation between GEBVs for the four SL measurements, GEBV of each individual for each measurement was calculated under the GBLUP model described in Chapter 1. Pearson's *r* between GEBVs of each SL measurement was computed using R/cor.test with default augments. Command line scripts are supplied as Appendix.

Results

The predictive ability for the SL_4 (harvest size, 569th day of the HY-treatment) was high (0.627–0.632) regardless of the prediction model (i.e., GBLUP, Bayes C, and Bayesian RKHS) (Table 5-3), confirming the feasibility of GS for SL of the individuals fed with HY-diet. Although body size at harvest is the primary target for selective breeding, evaluation at harvest is often time- and cost-consuming. If the harvest size could be predicted from the data at early stage, experimental time, feeding and other running costs can be dramatically decreased. To test the possibility of GP with a short rearing period, I investigated the predictive abilities of GP at harvest and three other time points. GEBV for SL_3 (422nd day) showed high correction between that for SL_4 (Pearson's r = 0.965), indicating the possibility of prediction for the harvest size from SL_3. The correlation was not even low between SL_1 (198th day) and SL_4 (0.915). Decreases in prediction accuracy of GP for the harvest size would not be negligible when the data of SL_1 was used. However, the economic gains from reduction in the operation time by one year may compensate the losses from the decreased prediction accuracy.

5.3 Exploit transcriptomic information for improvement of predictive ability

In the previous section, moderate predictive abilities of GP for SL of the individuals subjected to a long-term feeding treatment. This could be improved by taking the advantage of transcriptomic information obtained in Chapter 3: feeding HY-diet changed expression patterns of the genes related to terpenoid backbone biosynthesis and steroid biosynthesis. In this section, I have tested if predictive ability can be improved by exploiting the transcriptomic information; predictive ability for SL_4 was estimated under a GBLUP model incorporating a weighted genomic relationship matrix, where the genomic variants around and within genes underling backbone biosynthesis and steroid biosynthesis were up weighted to calculate genomic relationship between each pair of individuals.

Materials and methods

Weighted genomic relationship matrix

The genomic relationship matrix (Endelman and Jannink, 2012) used in the other Chapter following as follows:

$$\boldsymbol{G} = \frac{(\boldsymbol{M} - \boldsymbol{P})(\boldsymbol{M} - \boldsymbol{P})'}{2\sum_1^m p_i(1 - p_i)},$$

where **M** is a $n \times m$ (i.e., number of individuals × number of SNP markers) matrix of genotype represented as the number of minor alleles centered to zero, i.e., -1, 0, and 1; **P** is a $n \times m$ matrix in which each element of column i is $2(p_i - 0.5)$, i.e., p_i is the minor allele frequency at marker i. This method assumes that all markers equally contribute to genetic variation while it violates the facts that there are some markers which forms linkage disequilibrium with the QTLs with relatively larger effects.

In this section, larger effects were assigned for selected loci using a weighted genomic relationship matrix constructed following the method reported in Tiezzi and Maltecca (2015) (Tiezzi and Maltecca, 2015) with modifications:

$$\boldsymbol{G} = (\boldsymbol{M} - \boldsymbol{P})\boldsymbol{D}(\boldsymbol{M} - \boldsymbol{P})',$$

where $\boldsymbol{M}$ and $\boldsymbol{P}$ are same as unweighted genomic relationship matrix; $\boldsymbol{D}$ is a diagonal matrix with principal diagonal elements determined as follows:

$$D_{ii} = w_j,$$

where w_j is a weight assigned value j to selected markers, and 1 otherwise. Therefore, this weighted matrix uses manual weights w_j to represent the unequal importance of each marker. I tested four weights for selected markers, $w_1 = 1$, $w_2 = 2$, $w_5 = 5$, and $w_{10} = 10$. In this study, the SNPs around (100 kbp upstream and 100 kbp downstream) and within genes related to terpenoid backbone biosynthesis and steroid biosynthesis were upweighted as weighted SNP. Command line scripts are supplied as Appendix.

GBLUP with weighted genomic relationship matrix

GBLUP was applied following Chapter 1 with modification; weighted genomic relationship matrix was used instead of unweighted one. The cross-validation scheme was applied for calculation of predictive ability of GP following Chapter 2. Command line scripts are supplied as Appendix.

Effects of SNPs underling genes related to terpenoid backbone biosynthesis and steroid biosynthesis

The marker effects of GP on SL_4 were investigated using Bayes C model following Chapter 1. Average marker effect of the SNPs harboring from the100 kbp upstream the 100 kbp downstream of each gene related to terpenoid backbone biosynthesis and steroid biosynthesis was calculated. Command line scripts are supplied as Appendix.

Results

To improve the performance of GP, the transcriptomic information was exploited using GBLUP with weighted genomic relationship matrix. Specifically, terpenoid backbone biosynthesis and steroid biosynthesis include 24 and 27 genes, respectively, and ten genes were shared between them. There were 437 SNPs around (100 kbp upstream and 100 kbp downstream) and within these genes. These SNPs were weighted with w_j = 1, 2, 5, 10 for selected markers to construct weighted genomic relationship matrices. Using GBLUP with these matrices, predictive abilities for SL_4 were the same as or slightly lower (w_1 = 0.627 ± 0.006, mean ± S.E.; w_2 = 0.627 ± 0.006; w_5 = 0.626 ± 0.006; w_{10} = 0.623 ± 0.006) than the results in Section 5.2 (0.627–0.632), suggesting minimal genetic differences in each gene relating these pathways among individuals. Moreover, the predictive ability was gradually but not significantly decreased as the weights became higher, suggesting those genes underling the two pathways can explain a small fraction of phenotypic variation as the average of marker effects (absolute value) for SNPs around and within these genes (0.0039 ± 0.0033, S.D.) was similar to that for all SNPs estimated under the Bayes C model (0.0036 ± 0.0034, S.D.).

Discussion

In this chapter, I have investigated the effect of long-term treatment with HY diet, of which 67.5% of fishmeal was replaced with yeast protein, on genomic prediction for SL at harvest. Additionally, I have evaluated genetic correlation between SL measured at harvest and three different time points to examine the possibility of predicting GEBV for SL at harvest using phenotype recorded at early phase of production. In all, these results showed feasibility of GS for SL at harvest of individuals experienced a long-term dietary treatment of HY. Moreover, GP using SL data at the early stage of production (nine months, shortened almost one year) can satisfyingly predict GEBV at harvest, implying cost efficiency of GS for body size of the tiger pufferfish reared with HY-diet can be dramatically improved.

For each SL_i, moderate heritability was observed (SL_1: 0.548 ± 0.050; SL_2: 0.609 ± 0.048; SL_3: 0.612 ± 0.048; SL_4: 0.613 ± 0.048, h_2 ± S.E.), and estimated heritability increased as SL measurement at a longer period of feeding. This indicates the environmental effects on SL at an early age (a short feeding duration) are large than those on SL at the elder stages (a long feeding duration), as the growth of younger fish is more sensitive to the environmental factors such as tank effects, managements, feeding conditions (Gao et al., 2022; Hosoya et al., 2014). The heritability estimated (SL_1: 0.548 ± 0.050, approximately nine-month-old fish) in this chapter was slightly lower than that in Chapter 4 (Pop-C: 0.620 ± 0.046, approximately six-month-old fish), but it still confirmed generality of moderate heritability and supported the high potential for genetic improvement on body size under the HY-diet.

The predictive ability of GP for the SL_4 (harvest size, 569th day of the HY-treatment) was high (0.627–0.632), confirming the feasibility of GS for SL of the individuals fed with HY-diet. In addition, prediction using SL at earlier phase of production (SL_1, nine months) could effectively predict GEBV for SL at harvest (SL_4) as strong genetic correlation (r_g = 0.915) was observed and GEBV for the SL at each time point also showed strong correlation (r = 0.932–0.966). Given the breeder's equation (Lush, J. L., 1937):

$$\Delta G = i\, r\, \sigma_A / L,$$

where ΔG is the genetic gain over time; i is selection intensity; r is accuracy of selection; σ_A is the genetic variance of target trait among tested fish; L is generation interval. Although prediction using SL at an early age would have lower selection accuracy than that at harvest, this decrease would be compensated by the reduced farming costs to keep a thousand of individuals as the training groups for an additional year. In addition, the saved farming costs could be invested to increase the number of fish for the training group for prediction and selection candidates, leading increases in i, r, and σ_A, and thus improvements in ΔG.

In Chapter 3, terpenoid backbone biosynthesis and steroid biosynthesis were differentially expressed in the HY-fed group compared to FM-fed group (standard diet). Therefore, it was expected that the genes underling these two pathways have larger effects on growth performance of HY-fed fish than the others. In this study, I assumed these two pathways largely contribute to the body size trait and exploited these transcriptomic information to enhance the performance of GP using GBLUP with weighted genomic relationship matrix, which allow regions near selected markers to have more genetic variation than other regions (Pérez and de los Campos, 2014; Tiezzi and Maltecca, 2015). The model with weighted genomic relationship matrix theoretically has better

performance of GP when the assumption that large- or moderate-effect QTLs are near the weighted (selected) markers holds true (Liu et al., 2020). However, predictive abilities for SL_4 were the same as or slightly lower (w_1 = 0.627 ± 0.006, mean ± S.E.; w_2 = 0.627 ± 0.006; w_5 = 0.626 ± 0.006; w_{10} = 0.623 ± 0.006) than the results in Section 5.2 (0.627–0.632), while higher weights for selected markers caused lower predictive ability. These results suggested the assumption for weighted genomic relationship matrix may fail in this case, and genes related to terpenoid backbone biosynthesis and steroid biosynthesis possibly do not have larger effects on body size at harvest compared to the others. This is supported by the result that average absolute marker effects of selected SNPs (0.0039 ± 0.0033, S.D.) is almost same as the those of all SNPs (0.0036 ± 0.0034, S.D.). Additionally, differential expression of these two pathways (mRNA transcripts) may not reflect the changes of proteins as post-transcriptional regulation could control it as well (Dahan et al., 2011; Martin et al., 2016). Taken together, the utility of the transcriptomic information for improvements in predictive ability for the trait was not recognized.

General discussion

Genetic basis of heterobothriosis resistance and body size under a low fish meal diet in the tiger pufferfish and its potential for genetic improvement

This study revealed the polygenetic nature of heterobothriosis resistance and body size of the tiger pufferfish fed with both a standard diet and a LFM diet containing high propotion of yeast proteins (HY). In each case, the estimated heritability for the two traits were high enough for genetic improvement by means of selective breeding. It also supported the idea that disease resistance and growth performance in aquaculture species are commonly polygenetic, and have moderate heritability (Gjedrem and Rye, 2018). Moreover, these results also suggested that genomic selection (GS) is the choice to take for selective breeding for these traits instead of marker assisted section (MAS) as GS is suitable for polygenic traits while MAS is the powerful tool for mono- and oligogenic traits (Hayes et al., 2010; Wakchaure and Ganguly, 2015).

The higher heritability for heterobothriosis resistance and body size under a treatment of HY-diet compared to those under a standard feeding condition indicated the genetic basis was reshaped possibly due to gene-environment interactions (G×E); genes have specific responses under a particular environment (diet in this case). Besides this study, genotype-diet interactions (i.e., G×E under a particular diet) in production traits have been observed when testing the LFM diets in aquaculture species, such as in amago salmon (Yamamoto et al., 2014), rainbow trout (Le Boucher et al., 2011; Pierce et al., 2008) and European sea bass (Le Boucher et al., 2013). These findings suggest that reshape of genetic basis of production traits due to replacements of diets commonly occur in aquaculture species. Phenotype formulation under the effect of G×E can be expressed as follows:

$$P_i = G + E_i + GEI_i ,$$

where P_i is phenotype given for the population under a specific environment *i*, e.g., $i = 1$ for a standard diet and $i = 2$ for a LFM diet ; G is a vector of shared genetic effects among different environments; E_i is environmental effects under a specific environment *i*; GEI_i is effects from G×E for environment *i* (Baye et al., 2011). The heritability can be expressed as

$$h_i^2 = \frac{var(G_i)}{var(P_i)} = \frac{var(G + GEI_i)}{var(G + E_i + GEI_i)} = \frac{var(G) + var(GEI_i)}{var(G) + var(E_i) + var(GEI_i)},$$

where G_i is total genetic effects including G and GEI_i under the environment *i*. Given the fact that broodstocks for aquaculture species feed animal proteins for generations but barely experienced alternative proteins such as from plants, yeast, or bacteria; $var(GEI_i)$ for LFM diets would be larger than that for the regular feed; not all individuals can utilize these protein sources. However, $var(GEI_i)$ is practically not realized but absorbed in $var(G)$ and/or error components (residuals), which is not modeled in the aquation above. On the other hand, the size of $var(E_i)$ is diet-specific, and thus estimated heritability becomes higher when the feed brings about lower $var(E_i)$. In the case of SL, $var(E_i)$ for HY-diet would have been smaller than that for the regular diet since mean SL was larger in the HY-fed group than the others while variation in SL was not different between the diet groups, resulting in relatively higher heritability, although other factors might have contributed for the higher heritability; e.g. the difference could be merely due to the differences in the population structure. Although the

heritability for the two traits were higher under the HY-diet treatment than those under a standard feeding condition, it is possible that a certain LFM diet treatment results in lower heritability for production traits especially when the diet increases $var(E_i)$, e.g., the diet which not all fish can digest.

Genetic improvement for heterobothriosis resistance and body size under a low fish meal diet treatment in the tiger pufferfish by means of genomic selection

Moderate to high predictive abilities of GP for heterobothriosis resistance and body size traits under a treatment of standard diet or HY-diet confirmed the feasibility of GS, and thus a high selection accuracy of elite candidates and genetic gain could be expected in a GS breeding program. The possibility of selective breeding targeting production traits of fishes reared with a LFM or no fishmeal diet was rarely tested in aquaculture species, but in amago salmon (Yamamoto et al., 2016, 2015, 2014) and rainbow trout (Callet et al., 2017; Overturf et al., 2013) based on family selection. In contrast, in this dissertation, it is the first study that verified the feasibility of genetic improvement for production traits of the fish fed with LMF diets by means of a state-of-the-art breeding technology, GS. Rather than plant proteins, this yeast-based protein showed high potential for feed ingredients, and thus it can contribute improvements in the sustainability of aquaculture.

HY-diet and other three diets are designed to have similar energy level and nutritional component with the standard diet to fit the digestive system of fish, and HY-fed fish outperformed. Theoretically, the GS breeding strategy for fishmeal tolerance was not limited to the best diet, it could be used for an unqualified diet that is not properly fit the digestive system of farmed carnivorous fish while it brings about huge advantages in cost-efficiency. As an empirical example, the domestication of dogs from wolfs started at least 10,000 years before present (Skoglund et al., 2011) showed the signal of selection on tolerance to a starch-rich diet (Arendt et al., 2016; Axelsson et al., 2013), which is indigestible for many carnivorous animals. Likewise, the genetic improvement of tolerance to harsh dietary change like domestication of dogs could be theoretically realized in carnivorous fish using a GS strategy in a much shorter time than e.g., the dog domestication, and thus it could give more opportunity to utilize novel fishmeal replacements in aquaculture species.

On the other side, the genetic breeding of the tiger pufferfish is moving forward surely. Specifically, the identification of sex determining locus, i.e., Y specific SNP (Kamiya et al., 2012), facilitated the production super-male (YY male) of which progenies are all males. This enabled production of all-male populations, favored by hatcheries since males have more values because testes of this species have high market value (Hamasaki et al., 2013). Recently, this technique has been integrated with the surrogate broodstock technology (Takeuchi et al., 2004; Yoshizaki and Yazawa, 2019), where Y eggs of the tiger pufferfish are produced from glass pufferfish (*Takifugu alboplumbeus*), and the efficiency for the production of the all-male population has been dramatically improved (Hamasaki et al., 2017). Furthermore, GS program targeting precocious maturation has launched (Hosoya et al., 2021; Yoshikawa et al., 2021, 2020). The results of this study strongly suggest the possibility of GS for LFM tolerate breeds can be the next success story of the genetic breeding of this species.

Improving cost-efficiency of genomic selection by next-generation sequencing

This study clearly indicated the feasibility of GS for production of LFM tolerate breeds. However, it is required for GS to genotype hundreds or thousands of individuals and capturing the substantial genome-wide genetic variance among individuals at each breeding cycle (Meuwissen and Goddard, 2001). Thus, genotyping cost can limit the utilization of GS to establish elite breeds with LFM tolerance. On the other hand, recent advancements in NGS-based sequencing technologies can effectively reduce genotyping cost and facilitate the use of GS in practice. For example, the utility of AmpliSeq and GRAS-Di as the genotyping tools for the tiger puffer fish is confirmed in this study and others (Hosoya et al., 2021; Yoshikawa et al., 2021).

NGS technologies, which can generate billions of bases of DNA sequence in parallel at low cost and in a much shorter time (Behjati and Tarpey, 2013), are now used as a major genotyping strategies for GS in aquaculture. The genotyping technologies used in this study, namely AmpliSeq and GRAS-Di, belongs to genotyping-by-sequencing (GBS) technologies. In GBS, a subset of the genome, instead of the whole-genome, is sequenced (i.e., reduced-representation sequencing, RRS), and the sequence information is directly applied for genotype calling (Baird et al., 2008; Elshire et al., 2011; Sato et al., 2019). More importantly, recent studies revealed that a several thousands of SNPs are sufficient for GP in the aquaculture species, which did not experience long-term artificial selection dislike major crops and livestock (Hosoya et al., 2021, 2017; Kriaridou et al., 2020). This idea was also supported by the result of my study that 5,000 SNPs were enough to gain prediction accuracy comparable to that obtain from more than 15K-SNPs (Chapter 2). The genotyping costs of AmpliSeq and GRAS-Di can be further reduced by intensive multiplexing (Sato et al., 2019)ss. AmpliSeq adopts target-amplicon sequencing strategy which is high flexible for designing the number of SNPs and samples per run. Meanwhile, GRAS-Di based on a random-amplicon sequencing and superior in cost efficiency as its sequence library is constructed by two steps of multiplexing PCR with limited number of primer sets (Hosoya et al., 2019). Considering the performance of these two NGS-based methods in this study and others, genotyping cost would not be the major obstacle to put GS in practice in aquaculture.

Boosting GS applications in other aquaculture species

Compared to the major crops and livestock, aquaculture have diverse group of finfish and shellfish (543 different species); meanwhile, domestication and selective breeding of aquaculture species is at an early stage, except salmonids and tilapia (Houston et al., 2020). It is anticipated that implementation of GS in the aquaculture species with systematic breeding program improve the farming efficiency and sustainability. In this study, a pilot study was first conducted to characterize the genetic architecture of the trait and examine the feasibility of GS under a standard diet, and then a middle-scale experiment was performed to study the generality of the pilot study. The diet evaluation was conducted to test the utility of LFM diet for GS. Finally, the feasibility of GS for production traits under the LFM diet treatment. This investigation scheme could be referred as a model strategy and applied to the other aquaculture species with modifications, especially, for the carnivorous species that consumed a large amount of fishmeal.

In all, this study uses the tiger pufferfish as a model species and systematically investigated the feasibility of GS for production traits of the individuals fed with LFM diet. With these empirical results and this GS strategy, it is highly expected to establish fish breeds with high LFM tolerance in the tiger pufferfish and the other aquaculture species. Taking advantage of these elite fish breeds, the aquaculture sector could be more productive in a sustainable way.

論文の内容の要旨

水圏生物科学専攻

平成３１年度 博士課程進学

氏名 林 子杰(Lin Zijie)

指導教員名 菊池 潔

論文題目 Genetic basis and genetic improvement of heterobothriosis resistance and growth performance in the tiger pufferfish, *Takifugu rubripes*, fed with low fishmeal diet
(低魚粉飼料給餌下におけるトラフグのヘテロボツリウム症耐性と成長能力に関する遺伝基盤と選抜育種の可能性に関する研究)

Fishmeal produced from wild-captured fish is an important component of diets used in the aquaculture industry. However, the rising price of fishmeal and the necessity for sustainable fishing have encouraged the aquaculture industry to explore alternative protein resources. To reduce the amount of fishmeal, a wide range of potential replacements has been tested mainly using plant proteins and recently the potential of single cell proteins from microorganisms such as bacteria, yeasts, and microalgae have been suggested. However, these novel diets are still under development and remain untested for the majority of aquaculture species.

The availability of low fishmeal (LFM) diet would be largely depending on species specific characteristics of digestive system, and LFM diets may diminish growth performances of farmed fish. Even though, it is expected that growth performances of fish fed with LFM diets can be improved by harnessing selective breeding technologies. Selective breeding is the process to selectively improve economically important traits in farmed populations by producing genetically superior progenies through recurrent mating among high-potential individuals and has dramatically improved farming efficiency of major crops and live stocks. In aquaculture, its potential has been widely recognized and routinely practiced in several species, mainly salmonids and tilapias. Recently, genomic selection (GS) is attracting the most attention as a rapid selective breeding method, which leverage the genomic information to predict the genetic ability for artificial selection. Thanks to the developments in the next-generation sequencing technologies, the genome-wide single nucleotide polymorphisms (SNPs) is available for aquaculture species at low costs, and the feasibility of GS has been proven in a wide range of aquaculture species.

In this study, I have investigated the availability of GS for parasitic resistance against heterobothriosis, caused by a monogenean parasite *Heterobothrium okamotoi*, and growth performances in the tiger pufferfish, *Takifugu rubripes*, reared under a LFM diet. The species is a major aquaculture species in Japan that commands a high market price. Currently, the standard feeds for the tiger pufferfish are composed of extruded dry pellets

(EP) that contain a high level of animal protein (70–80%) mainly from fishmeal; thus, it is expected that the combination of the use of low fishmeal diets and GS can significantly reduce feed costs while facilitating sustainable aquaculture.

Chapter 1. Availability of genomic selection for heterobothriosis resistance and body size under a standard feed

I have first investigated the genetic basis of heterobothriosis resistance and growth trait, i.e., standard length (SL), and feasibility of GS for these traits using a small number of samples. The test population was produced by artificial mating among 11 wild males and 10 wild females. At four months old, 240 individuals were subjected to an artificial infection experiment. At 32 days post infection, SL were measured (mean ± S.D. = 9.83 ± 0.78, cm) and the number of parasites on branchial cavity walls (*H. okamotoi* counts, HC: 15.85 ± 9.15) was counted for each specimen. Subsequently, genome wide 6,707 SNPs were genotyped for all individuals. Genome-wide association study (GWAS) showed no SNPs significantly associated with each trait, implying these traits are polygenic. Genomic heritability (h^2, mean ± S.E.) estimated under genomic best linear unbiassed predictor (GBLUP) model was moderate for each of the traits (HC: 0.308 ± 0.123; SL: 0.405 ± 0.131), suggesting the high potential for genetic improvement by GS.

Feasibility of GS can be evaluated from predictive ability of the GS model. I have calculated prediction ability under 12 models, including GBLUP, Bayesian regressions (Bayes A, Bayes B, Bayes C, Bayesian ridge regression, Bayesian LASSO, and Bayesian RKHS), and machine learning procedures (feedforward neural networks and multi-task feedforward deep neural network). Moderate predictive ability was observed in each model for both HC (0.248–0.344) and SL (0.340–0.481), confirming the feasibility of GS for both traits. Notably, undesirable genetic correlation was observed between HC and SL (r_g = 0.228) under a multivariate GBLUP model. To examine the possibility of simultaneous improvements in both traits, in silico selection was performed for ten generations under six scenarios: random mating, selection on HC, selection on SL, simultaneous selection based on genomic Smith-Hazel index, and that based on the desired gains index. The simulation results suggested that selection on one trait may have negative response on the other and the desired gains index can help achieving simultaneous genetic improvements in the two traits.

Chapter 2. Genomic prediction for heterobothriosis resistance and body size traits in a middle-scale population

As described above, the feasibility of GS on heterobothriosis resistance and body size was confirmed using a small population produced from wild parents at a small-scale experiment. In this chapter, I have further tested the availability of GS using a middle-scale population produced using commercially raised individuals (ten males and four females) derived from families cultured for several generation. I have also tested the effects of the number of SNPs on the predictive ability. A total of 1,100 fish at four months old were subjected to an artificial infection trial to collect phenotypes of HC (3.31 ± 3.26) and SL (12.97 ± 0.95, cm). Genotypes of 12,548 SNPs were obtained for each fish after imputation. As in the case of the small-scale experiment, GWAS study revealed polygenetic nature of both traits. HC and SL showed low and moderate heritability, respectively (HC: h^2 = 0.162 ± 0.041; SL: h^2 = 0.537 ± 0.045). The predictive abilities estimated under three models, GBLUP, Bayes C and Bayesian RKHS, were also low for HC (0.253–0.258) and moderate for SL (0.520–0.524). The

low heritability and predictive ability for HC seen in the middle-scale experiment could be largely due to the phenotypic distribution leans to zero; nevertheless, the values still support the possibility of GS for heterobothriosis resistance. To examine the needed SNP density for GP, a panel of SNP density (12,548, 10,000, 7500, 5,000, 2,500, 1,000, 500, 100, 50 SNPs) were randomly sampled for 50 times to calculate predictive abilities. The results showed that average predictive abilities (± S.E.M.) of GP for HC using 5,000 SNPs (0.256 ± 0.001) are comparable to the one using 12,548 SNPs (0.258 ± 0.000) but decreased as the number of SNPs was less than 5,000.

Chapter 3. Availability of low fishmeal diets for the tiger pufferfish

Exploitation of alternative protein sources for LFM diets has been highly demanded for the tiger pufferfish. In this study, I have investigated the availability of single cell proteins such as yeast and bacteria as replacement of fishmeal as no systematic study on the practicality of using single cell proteins in this species even though the high potential of these proteins has been realized in several fish species.

Growth performance, blood chemistry, resistance to heterobothriosis, and transcriptomic responses in the liver were compared between groups fed either a standard diet (FM, of which fishmeal ingredients occupied 80% of the total mass) or one of the four LFM diets; in three diets, 37.5% of the fishmeal (30% of the total mass) was replaced by plant (PP), bacterial (BAC), or yeast (LY)-derived ingredients; in the fourth, 62.5% (50% of the total mass) was replaced by yeast (HY). After a three-month feeding trial, the HY-fed fish was significantly larger in SL ($p = 0.031$) and body weight ($p = 0.008$) compared to control fish but lower in hematocrit value (Ht, $p = 0.009$). Mean total cholesterol levels in BAC-, LY- and HY-fed fish were lower than in control fish ($p = 0.015$, 0.007, and 0.047, respectively) while no significant difference was found for triglyceride, total protein, and glucose levels in plasma among the groups. Heterobothriosis resistance was also not different among diet groups, and the difference in Ht between HY and the other groups was diminished at the end of the 36-day infection trial. These analyses indicate that a LFM diet in which 67.5% of fishmeal is replaced by yeast protein (i.e., containing 50% of yeast protein in total mass) may be a suitable replacement for the commercial FM diet as it promotes growth but does not induce severe deleterious effects. The gene expression analyses in the liver showed up-regulation in genes on the terpenoid backbone biosynthesis and steroid synthesis pathways in LFM diet groups suggesting the importance of these pathway to be adjusted themselves to LFM diets for the species.

Chapter 4. Genomic prediction for heterobothriosis resistance and body size under a short-term treatment of low fish meal diet

As the availability of the LFM diet including high proportion of yeast protein (HY) was demonstrated selected as the best diet after the evaluation, I have examined heritability and the possibility of genomic selection in two different populations: Pop-C, consisting of 993 individuals (three-month old) produced from the commercial broodfish, and Pop-CW, 1,017 individuals (three-month old) produced from commercial broodfish and wild individuals). After 99 days (Pop-C) and 128 days (Pop-CW) of dietary treatments with HY diet, fish were subjected to an artificial infection trial. These specimens were subsequently genotyped by means of GRAS-Di, resulting 19,227 and 21,272 imputed SNPs for Pop-C and Pop-CW, respectively. The predictive

abilities estimated under three models, i.e., GBLUP, Bayes C and Bayesian RKHS, were high for both HC (Pop-C: 0.520–0.522; Pop-CW: 0.564–0.567) and SL (Pop-C: 0.601–0.604; Pop-CW: 0.736–0.740), indicating the feasibility of GS for parasite resistance and growth performance under short-term HY diet feeding. Meanwhile, a strong genetic correlation between HC and SL were observed in both populations (r_g = 0.661 and 0.585, respectively), and therefore breeders need to pay close attention to this undesired correlation for selection.

Chapter 5. Genomic prediction for body size under a long-term treatment of low fish meal diets

I have demonstrated the feasibility of genomic prediction for heterobothriosis resistance and body size of the tiger pufferfish fed with HY-diet for a short period of several months. However, a long-term treatment of HY-diet may result in different outcomes. In this chapter, the ability of genetic prediction was investigated using a total of 936 three-month old fish, sharing the same parental individuals with Pop-C in Chapter 4. These specimens were fed with HY-diet for 19 months from juvenile (three months old) to the harvest size, during which standard length was recorded four times: SL_1 (198th day of the long-term feeding treatment), SL_2 (319th day), SL_3 (422nd day), and SL_4 (569th day). A total of 16,471 imputed SNPs were obtained. None of these SNPs showed significant association with SL at each sampling point. High genetic correlation between each pair of SL_i (r_g = 0.917–0.973) suggested high propotion of shared genetic architecture. Predictive abilities accessed by Bayesian RKHS were moderate and similar between SL at each sampling point (0.589–0.632), and GEBVs showed high correlation between each pair of SL_i (Pearson's r = 0.915–0.962) suggested harvest size can be predicted at early stage under HY-diet feeding condition. As the importance of genes related to terpenoid backbone biosynthesis and steroid biosynthesis for utilization of LFM diet was suggested in the gene expression analysis, it is expected that this transcriptomic information can be utilized to improve predictive ability. The predictive ability for SL_4 was estimated under the GBLUP model using weighted genomic relationship matrix (i.e., weighted SNPs near and within the genes in terpenoid backbone biosynthesis and steroid biosynthesis); however, predictive ability was lower when weighted genomic relationship matrix was used (0.623–0.627), suggesting minimal genetic differences in genes relating these pathways among individuals.

In conclusion, I found that both heterobothriosis resistance and SL are polygenic traits in the tiger pufferfish populations fed with the standard diet and the HY based LFM diet and demonstrated the feasibility of GS for these traits in the populations fed with either of the diets. It is expected that these findings and investigation strategies can be applied for the establishment of breeding populations with a high tolerance of low fishmeal diets, and thus contribute facilitating sustainable aquaculture of the tiger pufferfish and the other aquaculture fishes.

Acknowledgements

I would first like to thank Professor Dr. Kiyoshi Kikuchi, Fisheries Laboratory, Graduate School of Agricultural and Life Sciences, University of Tokyo, for his helpful guidance and continued encouragement. I would like to express my sincere appreciation to Dr. Sho Hosoya for his excellent teaching and I am gratefully indebted to his very valuable comments on this dissertation.

I am thankful to Mr. Naoki Mizuno for his efforts on management of fish. I am thankful to Dr. Mana Sato for the consultation of AmpliSeq. I thank to Dr. Takuya Itou and Dr. Yuki Kobayashi for their supports for running Illumina MiSeq sequencing. I thank Dr. Sota Yoshikawa and Dr. Masaomi Hamasaki, Nagasaki Prefectural Institute of Fisheries for their efforts on management of experimental specimens and experiment conduction. I am thankful to Dr. Takashi Koyama for the implementation of RNA-sequencing. I also thank all other members in my laboratory for their kind help.

Finally, I would like to thank my family: my parents, Qing Miao and Yufeng Lin, for their spiritual and financial support; my wife, Beibei Wang, for her love, companionship, and caring; my daughter, Yumu Lin, who always cheers me up.

Reference

Aas, T.S., Grisdale-Helland, B., Terjesen, B.F., Helland, S.J., 2006. Improved growth and nutrient utilisation in Atlantic salmon (*Salmo salar*) fed diets containing a bacterial protein meal. Aquaculture, 259, 365-376. https://doi.org/10.1016/j.aquaculture.2006.05.032

Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., Corrado, G.S., Davis, A., Dean, J., Devin, M., Ghemawat, S., Goodfellow, I., Harp, A., Irving, G., Isard, M., Jia, Y., Jozefowicz, R., Kaiser, L., Kudlur, M., Levenberg, J., Mane, D., Monga, R., Moore, S., Murray, D., Olah, C., Schuster, M., Shlens, J., Steiner, B., Sutskever, I., Talwar, K., Tucker, P., Vanhoucke, V., Vasudevan, V., Viegas, F., Vinyals, O., Warden, P., Wattenberg, M., Wicke, M., Yu, Y., Zheng, X., 2016. TensorFlow: Large-scale machine learning on heterogeneous distributed systems. In Proceedings of the 12th USENIX conference on Operating Systems Design and Implementation (OSDI'16). USENIX Association, USA, 265–283

Adelizi, P.D., Rosati, R.R., Warner, K., Wu, Y. V., Muench, T.R., White, M.R., Brown, P.B., 1998. Evaluation of fish-meal free diets for rainbow trout, *Oncorhynchus mykiss*. Aquac. Nutr., 4, 255–262. https://doi.org/10.1046/j.1365-2095.1998.00077.x

Agboola, J.O., Øverland, M., Skrede, A., Hansen, J.Ø., 2021. Yeast as major protein-rich ingredient in aquafeeds: a review of the implications for aquaculture production. Rev. Aquac., 13, 949–970. https://doi.org/10.1111/raq.12507

Alpern, D., Gardeux, V., Russeil, J., Mangeat, B., Meireles-Filho, A.C.A., Breysse, R., Hacker, D., Deplancke, B., 2019. BRB-seq: Ultra-affordable high-throughput transcriptomics enabled by bulk RNA barcoding and sequencing. Genome Biol., 20, 71. https://doi.org/10.1186/s13059-019-1671-x

Amir, E.A.D., Davis, K.L., Tadmor, M.D., Simonds, E.F., Levine, J.H., Bendall, S.C., Shenfeld, D.K., Krishnaswamy, S., Nolan, G.P., Pe'Er, D., 2013. viSNE enables visualization of high dimensional single-cell data and reveals phenotypic heterogeneity of leukemia. Nat. Biotechnol., 31, 545–552. https://doi.org/10.1038/nbt.2594

Arendt, M., Cairns, K.M., Ballard, J.W.O., Savolainen, P., Axelsson, E., 2016. Diet adaptation in dog reflects spread of prehistoric agriculture. Heredity (Edinb), 117(5), 301–306. https://doi.org/10.1038/hdy.2016.48

Arruda, M.P., Lipka, A.E., Brown, P.J., Krill, A.M., Thurber, C., Brown-Guedira, G., Dong, Y., Foresman, B.J., Kolb, F.L., 2016. Comparing genomic selection and marker-assisted selection for Fusarium head blight resistance in wheat (*Triticum aestivum L.*). Mol. Breed., 36, 84. https://doi.org/10.1007/s11032-016-0508-5

Asahida, T., Kobayashi, T., Saitoh, K., Nakayama, I., 1996. Tissue preservation and total DNA extraction from fish stored at ambient temperature using buffers containing high concentration of urea. Fish. Sci., 62, 727–730. https://doi.org/10.2331/fishsci.62.727

Awad, M., Khanna, R., Awad, M., Khanna, R., 2015. Support vector regression, in: Efficient Learning Machines. Apress, 67–80.

Axelsson, E., Ratnakumar, A., Arendt, M.L., Maqbool, K., Webster, M.T., Perloski, M., Liberg, O., Arnemo, J.M., Hedhammar, Å., Lindblad-Toh, K., 2013. The genomic signature of dog domestication reveals adaptation to a starch-rich diet. Nature, 495, 360–364. https://doi.org/10.1038/nature11837

Azodi, C.B., Bolger, E., McCarren, A., Roantree, M., de los Campos, G., Shiu, S.H., 2019. Benchmarking parametric and machine learning models for genomic prediction of complex traits. G3 Genes, Genomes, Genet., 9, 3691–3702. https://doi.org/10.1534/g3.119.400498

Baird, N.A., Etter, P.D., Atwood, T.S., Currey, M.C., Shiver, A.L., Lewis, Z.A., Selker, E.U., Cresko, W.A., Johnson, E.A., 2008. Rapid SNP discovery and genetic mapping using sequenced RAD markers. PLoS One, 3(10), e3376. https://doi.org/10.1371/journal.pone.0003376

Bangera, R., Ødegård, J., Præbel, A.K., Mortensen, A., Nielsen, H.M., 2011. Genetic correlations between growth rate and resistance to vibriosis and viral nervous necrosis in Atlantic cod (*Gadus morhua L.*). Aquaculture, 317, 67–73. https://doi.org/10.1016/j.aquaculture.2011.04.018

Baye, T.M., Abebe, T., Wilke, R.A., 2011. Genotype-environment interactions and their translational implications. Per. Med., 8(1), 59-70. https://doi.org/10.2217/pme.10.75

Behjati, S., Tarpey, P.S., 2013. What is next generation sequencing? Arch. Dis. Child. Educ. Pract. Ed., 98(6), 236-8. https://doi.org/10.1136/archdischild-2013-304340

Belghit, I., Waagbø, R., Lock, E.-J., Liland, N.S., 2019. Insect-based diets high in lauric acid reduce liver lipids in freshwater Atlantic salmon. Aquac. Nutr., 25, 343–357. https://doi.org/10.1111/anu.12860

Bentsen, H.B., Olesen, I., 2002. Designing aquaculture mass selection programs to avoid high inbreeding rates. Aquaculture, 204, 349-359. https://doi.org/10.1016/S0044-8486(01)00846-8

Biswas, A., Takakuwa, F., Yamada, S., Matsuda, A., Saville, R.M., LeBlanc, A., Silverman, J.A., Sato, N., Tanaka, H., 2020. Methanotroph (*Methylococcus capsulatus*, Bath) bacteria meal as an alternative protein source for Japanese yellowtail, *Seriola quinqueradiata*. Aquaculture, 529, 735700. https://doi.org/10.1016/j.aquaculture.2020.735700

Bolger, A.M., Lohse, M., Usadel, B., 2014. Trimmomatic: A flexible trimmer for Illumina sequence data. Bioinformatics, 30, 2114–2120. https://doi.org/10.1093/bioinformatics/btu170

Box, G.E.P., Cox, D.R., 1964. An Analysis of Transformations. J. R. Stat. Soc. Ser. B., 26(2), 211-252. https://doi.org/10.1111/j.2517-6161.1964.tb00553.x

Buetler, T.M., Krauskopf, A., Ruegg, U.T., 2004. Role of superoxide as a signaling molecule. News Physiol. Sci., 19(3), 120-123. https://doi.org/10.1152/nips.01514.2003

Caballero-Solares, A., Xue, X., Parrish, C.C., Foroutani, M.B., Taylor, R.G., Rise, M.L., 2018a. Changes in the liver transcriptome of farmed Atlantic salmon (*Salmo salar*) fed experimental diets based on terrestrial alternatives to fish meal and fish oil. BMC Genomics, 19(1), 796. https://doi.org/10.1186/s12864-018-5188-6

Callet, T., Médale, F., Larroquet, L., Surget, A., Aguirre, P., Kerneis, T., Labbé, L., Quillet, E., Geurden, I., Skiba-Cassy, S., Dupont-Nivet, M., 2017. Successful selection of rainbow trout (*Oncorhynchus mykiss*) on their ability to grow with a diet completely devoid of fishmeal and fish oil, and correlated changes in nutritional traits. PLoS One, 12(10), e0186705. https://doi.org/10.1371/journal.pone.0186705

Caruana, R., 1997. Multitask learning. Mach. Learn, 28, 41–75. https://doi.org/10.1023/A:1007379606734

Castillo-Juárez, H., Casares, J.C.Q., Campos-Montes, G., Villela, C.C., Ortega, A.M., Montaldo, H.H., 2007. Heritability for body weight at harvest size in the Pacific white shrimp, *Penaeus (Litopenaeus) vannamei*, from a multi-environment experiment using univariate and multivariate animal models. Aquaculture, 273(1), 42-49. https://doi.org/10.1016/j.aquaculture.2007.09.023

Cerón-Rojas, J.J., Crossa, J., 2018. Linear Selection Indices in Modern Plant Breeding. Springer.

Ceron-Rojas, J.J., Crossa, J., Arief, V.N., Basford, K., Rutkoski, J., Jarquín, D., Alvarado, G., Beyene, Y., Semagn, K., DeLacy, I., 2015. A genomic selection index applied to simulated and real data. G3 Genes, Genomes, Genet., 5, 2155–2164. https://doi.org/10.1534/g3.115.019869

Chen, Y., Chi, S., Zhang, S., Dong, X., Yang, Q., Liu, H., Zhang, W., Deng, J., Tan, B., Xie, S., 2021. Replacement of fish meal with Methanotroph (*Methylococcus capsulatus*, Bath) bacteria meal in the diets of Pacific white shrimp (*Litopenaeus vannamei*). Aquaculture, 541, 736801. https://doi.org/10.1016/j.aquaculture.2021.736801

Chen, Y., Lun, A.T.L., Smyth, G.K., 2016. From reads to genes to pathways: Differential expression analysis of RNA-Seq experiments using Rsubread and the edgeR quasi-likelihood pipeline. F1000Research, 5, 1438. https://doi.org/10.12688/F1000RESEARCH.8987.2

Chigasaki, M., Nakane, M., Ogawa, K., Wakabayashi, H., 2000. Standardized method for experimental infection of tiger puffer Takifugu rubripes with oncomiracidia of *Heterobothrium okamotoi* (Monogenea: Diclidophoridae) with some data on the oncomiracidial biology. Fish Pathol., 35, 215–221. https://doi.org/10.3147/jsfp.35.215

Chollet, F., others, 2015. keras.

Cooper, G.M., 2000. Peroxisomes, The Cell: A Molecular Approach. 2nd edition. Sinauer Associates, Sunderland (MA).

Cottrell, R.S., Blanchard, J.L., Halpern, B.S., Metian, M., Froehlich, H.E., 2020. Global adoption of novel aquaculture feeds could substantially reduce forage fish demand by 2030. Nat. Food., 1, 301–308. https://doi.org/10.1038/s43016-020-0078-x

Couture, J.L., Geyer, R., Hansen, J.Ø., Kuczenski, B., Øverland, M., Palazzo, J., Sahlmann, C., Lenihan, H., 2019. Environmental Benefits of Novel Nonhuman Food Inputs to Salmon Feeds. Environ. Sci. Technol., 53, 1967–1975. https://doi.org/10.1021/acs.est.8b03832

Covarrubias-Pazaran, G., 2018. Software update: Moving the R package sommer to multivariate mixed models for genome-assisted prediction. bioRxiv, 354639. https://doi.org/10.1101/354639

Covarrubias-Pazaran, G., 2016. Genome-assisted prediction of quantitative traits using the R package sommer. PLoS One, 11(6), e0156744. https://doi.org/10.1371/journal.pone.0156744

Crossa, J., 2012. From genotype × environment interaction to gene × environment interaction. Curr. Genomics, 13(3), 225-44. https://doi.org/10.2174/138920212800543066

Crossa, J., Pérez-Rodríguez, P., Cuevas, J., Montesinos-López, O., Jarquín, D., de los Campos, G., Burgueño, J., González-Camacho, J.M., Pérez-Elizalde, S., Beyene, Y., Dreisigacker, S., Singh, R., Zhang, X., Gowda, M., Roorkiwal, M., Rutkoski, J., Varshney, R.K., 2017. Genomic Selection in Plant Breeding: Methods, Models, and Perspectives. Trends Plant Sci., 22(11), 961-975. https://doi.org/10.1016/j.tplants.2017.08.011

Daetwyler, H.D., Calus, M.P.L., Pong-Wong, R., de los Campos, G., Hickey, J.M., 2013. Genomic prediction in animals and plants: Simulation of data, validation, reporting, and benchmarking. Genetics, 193(2), 347–365. https://doi.org/10.1534/genetics.112.147983

Daetwyler, H.D., Pong-Wong, R., Villanueva, B., Woolliams, J.A., 2010. The impact of genetic architecture on genome-wide evaluation methods. Genetics, 185, 1021–31. https://doi.org/10.1534/genetics.110.116855

Dahan, O., Gingold, H., Pilpel, Y., 2011. Regulatory mechanisms and networks couple the different phases of gene expression. Trends Genet., 27(8), 316-22. https://doi.org/10.1016/j.tig.2011.05.008

Dandine-Roulland, C., Bellenguez, C., Debette, S., Amouyel, P., Génin, E., Perdry, H., 2016. Accuracy of heritability estimations in presence of hidden population stratification. Sci. Rep., 6, 26471. https://doi.org/10.1038/srep26471

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., McVean, G., Durbin, R., 1000 Genomes Project Analysis Group, 1000 Genomes Project Analysis, 2011. The variant call format and VCFtools. Bioinformatics, 27, 2156–8. https://doi.org/10.1093/bioinformatics/btr330

Davidson, J., Barrows, F.T., Kenney, P.B., Good, C., Schroyer, K., Summerfelt, S.T., 2016. Effects of feeding a fishmeal-free versus a fishmeal-based diet on post-smolt Atlantic salmon Salmo salar performance, water quality, and waste production in recirculation aquaculture systems. Aquac. Eng., 74, 38–51. https://doi.org/10.1016/j.aquaeng.2016.05.004

De Los Campos, G., Gianola, D., Rosa, G.J.M., Weigel, K.A., Crossa, J., 2010. Semi-parametric genomic-enabled prediction of genetic

values using reproducing kernel Hilbert spaces methods. Genet. Res. (Camb)., 92, 295–308. https://doi.org/10.1017/S0016672310000285

De los Campos, G., Sorensen, D., Gianola, D., 2015. Genomic Heritability: What Is It? PLoS Genet., 11, e1005048. https://doi.org/10.1371/journal.pgen.1005048

Deng, J., Bi, B., Kang, B., Kong, L., Wang, Q., Zhang, X., 2013. Improving the growth performance and cholesterol metabolism of rainbow trout (*Oncorhynchus mykiss*) fed soyabean meal-based diets using dietary cholesterol supplementation. Br. J. Nutr., 110(1), 29-39. https://doi.org/10.1017/S0007114512004680

Desta, Z.A., Ortiz, R., 2014. Genomic selection: Genome-wide prediction in plant improvement. Trends Plant Sci., 19(9), pp., 592-601. https://doi.org/10.1016/j.tplants.2014.05.006

Egerton, S., Wan, A., Murphy, K., Collins, F., Ahern, G., Sugrue, I., Busca, K., Egan, F., Muller, N., Whooley, J., McGinnity, P., Culloty, S., Ross, R.P., Stanton, C., 2020. Replacing fishmeal with plant protein in Atlantic salmon (*Salmo salar*) diets by supplementation with fish protein hydrolysate. Sci. Rep., 10, 1–16. https://doi.org/10.1038/s41598-020-60325-7

Elaswad, A., Dunham, R., 2018. Disease reduction in aquaculture with genetic and genomic technology: current and future approaches. Rev. Aquac., 10, 876–898. https://doi.org/10.1111/raq.12205

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(5), e19379. https://doi.org/10.1371/journal.pone.0019379

Endelman, J.B., 2011. Ridge regression and other kernels for genomic selection with R package rrBLUP. Plant Genome J., 4, 250. https://doi.org/10.3835/plantgenome2011.08.0024

Endelman, J.B., Jannink, J.-L., 2012. Shrinkage estimation of the realized relationship matrix. G3 Genes, Genomes, Genet., 2, 1405–1413. https://doi.org/10.1534/g3.112.004259

Eng, S.K., Pusparajah, P., Ab Mutalib, N.S., Ser, H.L., Chan, K.G., Lee, L.H., 2015. Salmonella: A review on pathogenesis, epidemiology and antibiotic resistance. Front. Life Sci., 8(3), 284-293. https://doi.org/10.1080/21553769.2015.1051243

Enoki, H., 2019. The Construction of Psedomolecules of a Commercial Strawberry by DeNovoMAGIC and New Genotyping Technology, GRAS-Di. Proceedings of the Plant and Animal genome conference XXVII. San Diego, CA. Retrieved from https ://pag.confex.com/pag/ xxvii/ meeti ngapp.cgi/Paper/ 37002

Enoki, H., Takeuchi, Y., 2018. New Genotyping Technology, GRAS-Di, Using Next Generation Sequencer. Proceedings of the Plant and Animal genome conference XXVI. San Diego, CA. Retrieved from https ://pag. confex.com/pag/xxvi/meeti ngapp.cgi/Paper/ 29067

Esmaeili, M., 2021. Blood Performance: A New Formula for Fish Growth and Health. Biology (Basel), 10, 1236. https://doi.org/10.3390/biology10121236

Evenhuis, J.P., Leeds, T.D., Marancik, D.P., Lapatra, S.E., Wiens, G.D., 2015. Rainbow trout (*Oncorhynchus mykiss*) resistance to columnaris disease is heritable and favorably correlated with bacterial cold water disease resistance. J. Anim. Sci., 93, 1546–1554. https://doi.org/10.2527/jas.2014-8566

Faux, A.M., Gorjanc, G., Gaynor, R.C., Battagin, M., Edwards, S.M., Wilson, D.L., Hearne, S.J., Gonen, S., Hickey, J.M., 2016. AlphaSim: Software for Breeding Program Simulation. Plant Genome, 9, 1–14. https://doi.org/10.3835/plantgenome2016.02.0013

Frana, P.L., 2003. Genetic Prehistory in Selective Breeding: A Prelude to Mendel. Agric. Hist., 77(4), 619-620. https://doi.org/10.1525/ah.2003.77.4.619

Friesen, J.A., Rodwell, V.W., 2004. The 3-hydroxy-3-methylglutaryl coenzyme-A (HMG-CoA) reductases. Genome Biol., 5, 248 https://doi.org/10.1186/gb-2004-5-11-248

Fuji, K., Hasegawa, O., Honda, K., Kumasaka, K., Sakamoto, T., Okamoto, N., 2007. Marker-assisted breeding of a lymphocystis disease-resistant Japanese flounder (*Paralichthys olivaceus*). Aquaculture, 272, 291–295. https://doi.org/10.1016/j.aquaculture.2007.07.210

Fujiwara, R., Yokoi, T., Nakajima, M., 2016. Structure and protein-protein interactions of human UDP-glucuronosyltransferases. Front. Pharmacol, 7. https://doi.org/10.3389/fphar.2016.00388

Gamboa-Delgado, J., Márquez-Reyes, J.M., 2018. Potential of microbial-derived nutrients for aquaculture development. Rev. Aquac. 10, 224–246. https://doi.org/10.1111/raq.12157

Gao, X.Q., Wang, X., Wang, X.Y., Li, H.X., Xu, L., Huang, B., Meng, X.S., Zhang, T., Chen, H. Bin, Xing, R., Liu, B.L., 2022. Effects of different feeding frequencies on the growth, plasma biochemical parameters, stress status, and gastric evacuation of juvenile tiger puffer fish (*Takifugu rubripes*). Aquaculture, 548(2), 737718. https://doi.org/10.1016/j.aquaculture.2021.737718

Gatlin, D.M., Barrows, F.T., Brown, P., Dabrowski, K., Gaylord, T.G., Hardy, R.W., Herman, E., Hu, G., Krogdahl, Å., Nelson, R., Overturf, K., Rust, M., Sealey, W., Skonberg, D., Souza, E.J., Stone, D., Wilson, R., Wurtele, E., 2007. Expanding the utilization of sustainable plant products in aquafeeds: A review. Aquac. Res., 38(6), 551-579. https://doi.org/10.1111/j.1365-2109.2007.01704.x

Gaynor, R.C., Gorjanc, G., Hickey, J.M., 2020. AlphaSimR: An R-package for Breeding Program Simulations. bioRxiv, 2020.08.10.245167. https://doi.org/10.1101/2020.08.10.245167

Geay, F., Ferraresso, S., Zambonino-Infante, J.L., Bargelloni, L., Quentel, C., Vandeputte, M., Kaushik, S., Cahu, C.L., Mazurais, D., 2011a. Effects of the total replacement of fish-based diet with plant-based diet on the hepatic transcriptome of two European sea bass (*Dicentrarchus labrax*) half-sibfamilies showing different growth rates with the plant-based diet. BMC Genomics, 12, 522.

https://doi.org/10.1186/1471-2164-12-522

Gianola, D., Hayes, B.J., Goddard, M.E., Sorensen, D., Calus, M.P.L., 2013. Priors in whole-genome regression: the bayesian alphabet returns. Genetics, 194, 573–96. https://doi.org/10.1534/genetics.113.151753

Gianola, D., Okut, H., Weigel, K.A., Rosa, G.J., 2011. Predicting complex quantitative traits with Bayesian neural networks: a case study with Jersey cows and wheat. BMC Genet., 12, 87. https://doi.org/10.1186/1471-2156-12-87

Gjedrem, T., Baranski, M., 2009. Selective Breeding in Aquaculture: An Introduction. Springer. https://doi.org/10.1007/978-90-481-2773-3

Gjedrem, T., Robinson, N., Rye, M., 2012. The importance of selective breeding in aquaculture to meet future demands for animal protein: A review. Aquaculture, 350–353, 117–129. https://doi.org/10.1016/J.AQUACULTURE.2012.04.008

Gjedrem, T., Rye, M., 2018. Selection response in fish and shellfish: a review. Rev. Aquac., 10, 168–179. https://doi.org/10.1111/raq.12154

Gjerde, B., Ødegård, J., Thorland, I., 2011. Estimates of genetic variation in the susceptibility of Atlantic salmon (*Salmo salar*) to the salmon louse *Lepeophtheirus salmonis*. Aquaculture, 314, 66–72. https://doi.org/10.1016/j.aquaculture.2011.01.026

Glencross, B.D., Booth, M., Allan, G.L., 2007. A feed is only as good as its ingredients - A review of ingredient evaluation strategies for aquaculture feeds. Aquac. Nutr., 13(1), 17-34. https://doi.org/10.1111/j.1365-2095.2007.00450.x

Goddard, M.E., Hayes, B.J., 2009. Mapping genes for complex traits in domestic animals and their use in breeding programmes. Nat. Rev. Genet., 10, 381–391. https://doi.org/10.1038/nrg2575

Gumus, E., Aydin, B., Kanyilmaz, M., 2016. Growth and feed utilization of goldfish (*Carassius auratus*) fed graded levels of brewers yeast (*Saccharomyces cerevisiae*). Iran. J. Fish. Sci., 15(3), 1124 -1133.

Guo, Z., Tucker, D.M., Basten, C.J., Gandhi, H., Ersoz, E., Guo, B., Xu, Z., Wang, D., Gay, G., 2014. The impact of population structure on genomic prediction in stratified populations. Theor. Appl. Genet., 127, 749–762. https://doi.org/10.1007/s00122-013-2255-x

Habier, D., Fernando, R.L., Kizilkaya, K., Garrick, D.J., 2011. Extension of the bayesian alphabet for genomic selection. BMC Bioinformatics, 12, 186. https://doi.org/10.1186/1471-2105-12-186

Hamasaki, M., Takeuchi, Y., Miyaki, K., Yoshizaki, G., 2013. Gonadal Development and Fertility of Triploid Grass Puffer *Takifugu niphobles* Induced by Cold Shock Treatment. Mar. Biotechnol., 15, 133–144. https://doi.org/10.1007/s10126-012-9470-3

Hamasaki, M., Takeuchi, Y., Yazawa, R., Yoshikawa, S., Kadomura, K., Yamada, T., Miyaki, K., Kikuchi, K., Yoshizaki, G., 2017. Production of Tiger Puffer *Takifugu rubripes* Offspring from Triploid Grass Puffer *Takifugu niphobles Parents*. Mar. Biotechnol., 19, 579–591. https://doi.org/10.1007/s10126-017-9777-1

Haraga, A., Ohlson, M.B., Miller, S.I., 2008. Salmonellae interplay with host cells. Nat. Rev. Microbiol., 6, 53-66. https://doi.org/10.1038/nrmicro1788

Hauptman, B.S., Barrows, F.T., Block, S.S., Gibson Gaylord, T., Paterson, J.A., Rawles, S.D., Sealey, W.M., 2014. Evaluation of grain distillers dried yeast as a fish meal substitute in practical-type diets of juvenile rainbow trout, *Oncorhynchus mykiss*. Aquaculture, 432, 7-12. https://doi.org/10.1016/j.aquaculture.2014.03.026

Hayashi, S., Fujiwara, S., Noguchi, T., 1989. Degradation of Uric Acid in Fish Liver Peroxisomes. J. Biol. Chem., 264(6), 3211-3215. https://doi.org/10.1016/s0021-9258(18)94053-6

Hayes, B.J., Pryce, J., Chamberlain, A.J., Bowman, P.J., Goddard, M.E., 2010. Genetic Architecture of Complex Traits and Accuracy of Genomic Prediction: Coat Colour, Milk-Fat Percentage, and Type in Holstein Cattle as Contrasting Model Traits. PLoS Genet., 6, e1001139. https://doi.org/10.1371/journal.pgen.1001139

Hazel, L.N., 1943. The genetic basis for constructing selection indexes. Genetics, 28, 476–90.

Henderson, C.R., 1976. A Simple Method for Computing the Inverse of a Numerator Relationship Matrix Used in Prediction of Breeding Values. Biometrics, 32, 69. https://doi.org/10.2307/2529339

Henderson, C.R., 1953. Estimation of Variance and Covariance Components, Biometrics, 9(2), 226-252.

Hernández, C., Olvera-Novoa, M.A., Aguilar-Vejar, K., González-Rodríguez, B., Abdo de la Parra, I., 2008. Partial replacement of fish meal by porcine meat meal in practical diets for Pacific white shrimp (*Litopenaeus vannamei*). Aquaculture, 277, 244–250. https://doi.org/10.1016/j.aquaculture.2008.02.016

Hosoya, S., Hirase, S., Kikuchi, K., Nanjo, K., Nakamura, Y., Kohno, H., Sano, M., 2019. Random PCR-based genotyping by sequencing technology GRAS-Di (genotyping by random amplicon sequencing, direct) reveals genetic structure of mangrove fishes. Mol. Ecol. Resour., 19, 1153–1163. https://doi.org/10.1111/1755-0998.13025

Hosoya, S., Kido, S., Hirabayashi, Y., Kai, W., Kinami, R., Yoshinaga, T., Ogawa, K., Suetake, H., Kikuchi, K., Suzuki, Y., 2013. Genomic regions of pufferfishes responsible for host specificity of a monogenean parasite, *Heterobothrium okamotoi*. Int. J. Parasitol., 43, 909–915. https://doi.org/10.1016/j.ijpara.2013.06.006

Hosoya, S., Kikuchi, K., Nagashima, H., Onodera, J., Sugimoto, K., Satoh, K., Matsuzaki, K., Yasugi, M., Nagano, A.J., Kumagayi, A., Ueda, K., Kurokawa, T., 2018. Assessment of genetic diversity in Coho salmon (*Oncorhynchus kisutch*) populations with no family records using ddRAD-seq. BMC Res. Notes, 11, 548. https://doi.org/10.1186/s13104-018-3663-4

Hosoya, S., Kikuchi, K., Nagashima, H., Onodera, J., Sugimoto, K., Satoh, K., Matsuzaki, K., Yasugi, M., Nagano, A.J., Kumagayi, A., Ueda, K., Kurokawa, T., 2017. Genomic Selection in Aquaculture Breeding Programs. Bull. Japan Fish. Res. Educ. Agency, 35–39.

https://doi.org/10.1002/9781118782392.ch21

Hosoya, S., Mizuno, N., Kikuchi, K., Kurokura, H., 2014. Rearing Takifugu rubripes larvae in communal tanks: paternal genetic contribution to survivability. Fish. Sci., 80, 1037–1043. https://doi.org/10.1007/s12562-014-0795-x

Hosoya, S., Yoshikawa, S., Sato, M., Kikuchi, K., 2021. Genomic prediction for testes weight of the tiger pufferfish, *Takifugu rubripes*, using medium to low density SNPs. Sci. Rep., 11, 1–10. https://doi.org/10.1038/s41598-021-99829-1

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., Davie, A., Robledo, D., 2020. Harnessing genomics to fast-track genetic improvement in aquaculture. Nat. Rev. Genet., 21, 389–409. https://doi.org/10.1038/s41576-020-0227-y

Hua, K., Cobcroft, J.M., Cole, A., Condon, K., Jerry, D.R., Mangott, A., Praeger, C., Vucko, M.J., Zeng, C., Zenger, K., Strugnell, J.M., 2019. The Future of Aquatic Protein: Implications for Protein Sources in Aquaculture Diets. One Earth, 1(3), 316-329. https://doi.org/10.1016/j.oneear.2019.10.018

Huyben, D., Vidakovic, A., Nyman, A., Langeland, M., Lundh, T., Kiessling, A., 2017. Effects of dietary yeast inclusion and acute stress on post-prandial whole blood profiles of dorsal aorta-cannulated rainbow trout. Fish Physiol. Biochem., 43, 421–434. https://doi.org/10.1007/s10695-016-0297-0

Ido, A., Ali, M.F.Z., Takahashi, T., Miura, C., Miura, T., 2021. Growth of yellowtail (*Seriola quinqueradiata*) fed on a diet including partially or completely defatted black soldier fly (*hermetia illucens*) larvae meal. Insects, 12(8), 722. https://doi.org/10.3390/insects12080722

Igarashi, K., Matsunaga, R., Hirakawa, S., Hosoya, S., Suetake, H., Kikuchi, K., Suzuki, Y., Nakamura, O., Miyadai, T., Tasumi, S., Tsutsui, S., 2017. Mucosal IgM antibody with d-Mannose affinity in fugu *Takifugu rubripes* is utilized by a Monogenean parasite *Heterobothrium okamotoi* for host recognition. J. Immunol., 198, 4107–4114. https://doi.org/10.4049/jimmunol.1601996

Itoh, Y., Yamada, Y., 1987. Comparisons of selection indices achieving predetermined proportional gains. Genet. Sel. Evol., 19, 69. https://doi.org/10.1186/1297-9686-19-1-69

Iwata, H., Minamikawa, M.F., Kajiya-Kanegae, H., Ishimori, M., Hayashi, T., 2016. Genomics-assisted breeding in fruit trees. Breed. Sci., 66(1), 100-115. https://doi.org/10.1270/jsbbs.66.100

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

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

Jepson, M.A., Collares-Buzato, C.B., Clark, M.A., Hirst, B.H., Simmons, N.L., 1995. Rapid disruption of epithelial barrier function by Salmonella typhimurium is associated with structural modification of intercellular junctions. Infect. Immun., 63(1), 356-359. https://doi.org/10.1128/iai.63.1.356-359.1995

Kai, W., Kikuchi, K., Tohari, S., Chew, A.K., Tay, A., Fujiwara, A., Hosoya, S., Suetake, H., Naruse, K., Brenner, S., Suzuki, Y., Venkatesh, B., 2011. Integration of the genetic map and genome assembly of fugu facilitates insights into distinct features of genome evolution in teleosts and mammals. Genome Biol. Evol., 3, 424–442. https://doi.org/10.1093/gbe/evr041

Kamiya, T., Kai, W., Tasumi, S., Oka, A., Matsunaga, T., Mizuno, N., Fujita, M., Suetake, H., Suzuki, S., Hosoya, S., Tohari, S., Brenner, S., Miyadai, T., Venkatesh, B., Suzuki, Y., Kikuchi, K., 2012. A trans-species missense SNP in Amhr2 is associated with sex determination in the tiger Pufferfish, Takifugu rubripes (Fugu). PLoS Genet., 8, e1002798. https://doi.org/10.1371/journal.pgen.1002798

Kanehisa, M., Goto, S., Furumichi, M., Tanabe, M., Hirakawa, M., 2009. KEGG for representation and analysis of molecular networks involving diseases and drugs. Nucleic Acids Res. 38, D355-D360. https://doi.org/10.1093/nar/gkp896

Karaman, E., Cheng, H., Firat, M.Z., Garrick, D.J., Fernando, R.L., 2016. An upper bound for accuracy of prediction using GBLUP. PLoS One, 11(8), e0161054. https://doi.org/10.1371/journal.pone.0161054

Kaushik, S.J., Covès, D., Dutto, G., Blanc, D., 2004. Almost total replacement of fish meal by plant protein sources in the diet of a marine teleost, the European seabass, *Dicentrarchus labrax*. Aquaculture, 230, 391–404. https://doi.org/10.1016/S0044-8486(03)00422-8

Kempthorne, O., Nordskog, A.W., 1959. Restricted Selection Indices. Biometrics, 15, 10. https://doi.org/10.2307/2527598

Kemski, M.M., Rappleye, C.A., Dabrowski, K., Bruno, R.S., Wick, M., 2020. Transcriptomic response to soybean meal-based diets as the first formulated feed in juvenile yellow perch (*Perca flavescens*). Sci. Rep., 10, 1–12. https://doi.org/10.1038/s41598-020-59691-z

Khosravi, S., Rahimnejad, S., Herault, M., Fournier, V., Lee, C.R., Dio Bui, H.T., Jeong, J.B., Lee, K.J., 2015. Effects of protein hydrolysates supplementation in low fish meal diets on growth performance, innate immunity and disease resistance of red sea bream *Pagrus major*. Fish Shellfish Immunol., 45, 858–868. https://doi.org/10.1016/j.fsi.2015.05.039

Kikuchi, K., Furuta, T., 2009. Use of defatted soybean meal and blue mussel meat as substitute for fish meal in the diet of tiger puffer, *takifugu rubripes*. J. World Aquac. Soc., 40, 472–482. https://doi.org/10.1111/j.1749-7345.2009.00265.x

Kikuchi, K., Honda, H., Kiyono, M., Miyazono, I., 1993. Total replacement of fish meal with other protein sources in the diet of Japanese flounder, *Paralichthys olivaceus*. Aquaculture Science., 41, 345–351. https://doi.org/10.11233/aquaculturesci1953.41.345

Kim, D.I., Hosoya, S., Mizuno, N., Ito, K., Ieda, R., Kikuchi, K., 2019. Genetic variation in resistance of the tiger pufferfish *Takifugu*

rubripes to a host-specific monogenean parasite *Heterobothrium okamotoi*. Fish. Sci., 85, 1019–1025. https://doi.org/10.1007/s12562-019-01342-y

Kitson, S.M., Mullen, W., Cogdell, R.J., Bill, R.M., Fraser, N.J., 2011. GPCR production in a novel yeast strain that makes cholesterol-like sterols. Methods, 55(4), 287-292. https://doi.org/10.1016/j.ymeth.2011.09.023

Kortner, T.M., Gu, J., Krogdahl, Å., Bakke, A.M., 2013. Transcriptional regulation of cholesterol and bile acid metabolism after dietary soyabean meal treatment in Atlantic salmon (*Salmo salar L.*). Br. J. Nutr., 109, 593–604. https://doi.org/10.1017/S0007114512002024

Kriaridou, C., Tsairidou, S., Houston, R.D., Robledo, D., 2020. Genomic Prediction Using Low Density Marker Panels in Aquaculture: Performance Across Species, Traits, and Genotyping Platforms. Front. Genet., 11, 124. https://doi.org/10.3389/fgene.2020.00124

Krueger, F., n.d. Babraham Bioinformatics.

Kuhn, M., 2008. Building predictive models in R using the caret package. J. Stat. Softw., 28(5), 1-26. https://doi.org/10.18637/jss.v028.i05

Lande, R., Thompson, R., 1990. Efficiency of marker assisted selection in the improvement of quantitative traits. Genetics 124, 743–756.

Law, C.W., Chen, Y., Shi, W., Smyth, G.K., 2014. Voom: Precision weights unlock linear model analysis tools for RNA-seq read counts. Genome Biol. 15, R29. https://doi.org/10.1186/gb-2014-15-2-r29

Lazzarotto, V., Mé Dale, F., Larroquet, L., Corraze, G., 2018. Long-term dietary replacement of fishmeal and fish oil in diets for rainbow trout (*Oncorhynchus mykiss*): Effects on growth, whole body fatty acids and intestinal and hepatic gene expression. PLoS One, 13(1), e0190730. https://doi.org/10.1371/journal.pone.0190730

Le Boucher, R., Quillet, E., Vandeputte, M., Lecalvez, J.M., Goardon, L., Chatain, B., Médale, F., Dupont-Nivet, M., 2011. Plant-based diet in rainbow trout (*Oncorhynchus mykiss Walbaum*): Are there genotype-diet interactions for main production traits when fish are fed marine vs. plant-based diets from the first meal? Aquaculture, 321, 41–48. https://doi.org/10.1016/j.aquaculture.2011.08.010

Le Boucher, R., Vandeputte, M., Dupont-Nivet, M., Quillet, E., Ruelle, F., Vergnet, A., Kaushik, S., Allamellou, J.M., Médale, F., Chatain, B., 2013. Genotype by diet interactions in European sea bass (*Dicentrarchus labrax L.*): Nutritional challenge with totally plant-based diets. J. Anim. Sci., 91(1), 44–56. https://doi.org/10.2527/jas.2012-5311

Leduc, A., Zatylny-Gaudin, C., Robert, M., Corre, E., Corguille, G. Le, Castel, H., Lefevre-Scelles, A., Fournier, V., Gisbert, E., Andree, K.B., Henry, J., 2018. Dietary aquaculture by-product hydrolysates: Impact on the transcriptomic response of the intestinal mucosa of European seabass (*Dicentrarchus labrax*) fed low fish meal diets. BMC Genomics, 19, 396. https://doi.org/10.1186/s12864-018-4780-0

Li, H., 2013. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv:1303.3997.

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. https://doi.org/10.1093/bioinformatics/btp352

Li, W., Cerise, J.E., Yang, Y., Han, H., 2017. Application of t-SNE to human genetic data. J. Bioinform. Comput. Biol., 15(4), 1750017. https://doi.org/10.1142/S0219720017500172

Lim, S.J., Kim, S.S., Ko, G.Y., Song, J.W., Oh, D.H., Kim, J.D., Kim, J.U., Lee, K.J., 2011. Fish meal replacement by soybean meal in diets for Tiger puffer, *Takifugu rubripes*. Aquaculture, 313, 165–170. https://doi.org/10.1016/j.aquaculture.2011.01.007

Lin, Z., Hosoya, S., Sato, M., Mizuno, N., Kobayashi, Y., Itou, T., Kikuchi, K., 2020. Genomic selection for heterobothriosis resistance concurrent with body size in the tiger pufferfish, *Takifugu rubripes*. Sci. Rep., 10, 19976. https://doi.org/10.1038/s41598-020-77069-z

Liu, A., Lund, M.S., Boichard, D., Karaman, E., Guldbrandtsen, B., Fritz, S., Aamand, G.P., Nielsen, U.S., Sahana, G., Wang, Y., Su, G., 2020. Weighted single-step genomic best linear unbiased prediction integrating variants selected from sequencing data by association and bioinformatics analyses. Genet. Sel. Evol., 52, 48. https://doi.org/10.1186/s12711-020-00568-0

Liu, H., Zhou, H., Wu, Y., Li, X., Zhao, J., Zuo, T., Zhang, X., Zhang, Y., Liu, S., Shen, Y., Lin, H., Zhang, Z., Huang, K., Lübberstedt, T., Pan, G., 2015. The impact of genetic relationship and linkage disequilibrium on genomic selection. PLoS One, 10, e0132379. https://doi.org/10.1371/journal.pone.0132379

Lynch, M., Walsh, B., 1998. Genetics and analysis of quantitative traits. Sinauer.

Lush, J. L. Animal Breeding Plans. Ames, Iowa: Iowa State Press, 1937.

Magalhães, R., Sánchez-López, A., Leal, R.S., Martínez-Llorens, S., Oliva-Teles, A., Peres, H., 2017. Black soldier fly (*Hermetia illucens*) pre-pupae meal as a fish meal replacement in diets for European seabass (*Dicentrarchus labrax*). Aquaculture, 476, 79–85. https://doi.org/10.1016/j.aquaculture.2017.04.021

Maita, M., Maekawa, J., Satoh, K.I., Futami, K., Satoh, S., 2006. Disease resistance and hypocholesterolemia in yellowtail *Seriola quinqueradiata* fed a non-fishmeal diet. Fish. Sci., 72, 513–519. https://doi.org/10.1111/j.1444-2906.2006.01179.x

Maleki, F., Ovens, K., Hogan, D.J., Kusalik, A.J., 2020. Gene set analysis: challenges, opportunities, and future research. Front. Genet., 11. https://doi.org/10.3389/fgene.2020.00654

Martin, M., 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet.journal, 17, 10. https://doi.org/10.14806/ej.17.1.200

Martin, S.A.M., Dehler, C.E., Król, E., 2016. Transcriptomic responses in the fish intestine. Dev. Comp. Immunol., 64, 103-117. https://doi.org/10.1016/j.dci.2016.03.014

Matassa, S., Boon, N., Pikaar, I., Verstraete, W., 2016. Microbial protein: future sustainable food supply route with low environmental footprint. Microb. Biotechnol., 9, 568–575. https://doi.org/10.1111/1751-7915.12369

Mathew, B., Léon, J., Sillanpää, M.J., 2018. A novel linkage-disequilibrium corrected genomic relationship matrix for SNP-heritability estimation and genomic prediction. Heredity (Edinb), 120, 356–368. https://doi.org/10.1038/s41437-017-0023-4

Mathur, R., Rotroff, D., Ma, J., Shojaie, A., Motsinger-Reif, A., 2018. Gene set analysis methods: A systematic comparison. BioData Min., 11, 8. https://doi.org/10.1186/s13040-018-0166-8

Matsui, S., Goto, T., Tsubouchi, Y., Hirakawa, S., Suetake, H., Miyadai, T., Nakamura, O., Tasumi, S., Tsutsui, S., 2020. D-mannose-specific immunoglobulin M in grass puffer (*Takifugu niphobles*), a nonhost fish of a monogenean ectoparasite *Heterobothrium okamotoi*, can act as a trigger for its parasitism. J. Parasitol., 106, 276. https://doi.org/10.1645/19-21

McCarthy, D.J., Smyth, G.K., 2009. Testing significance relative to a fold-change threshold is a TREAT. Bioinformatics, 25, 765–771. https://doi.org/10.1093/bioinformatics/btp053

McDermaid, A., Monier, B., Zhao, J., Liu, B., Ma, Q., 2019. Interpretation of differential gene expression results of RNA-seq data: Review and integration. Brief. Bioinform. 20(6), 2044–2054. https://doi.org/10.1093/bib/bby067

Meuwissen, T.H., Goddard, M.E., 2001. Prediction of identity by descent probabilities from marker-haplotypes. Genet. Sel. Evol., 33, 605–634. https://doi.org/10.1186/1297-9686-33-6-605

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

Miura, M., Yamamoto, T., Ozawa, R., Okazaki, T., Murashita, K., Oku, H., Matsunari, H., Furuita, H., Mano, N., Suzuki, N., 2020. A preliminary study toward the improvement of low fishmeal diet utilization in a Yamanashi strain of rainbow trout *Oncorhynchus mykiss*. Aquac. Sci., 67(2), 127-138. https://doi.org/10.11233/aquaculturesci.67.127

Miyaki, K., Tyuda, H., Watanabe, T., Mizuta, K., Tsukashima, Y., Yoshida, N., TABETA, O., 1998. Treatment of tiger puffer, *takifugu rubripes*, eggs with tannic acid to eliminate their adhesiveness for seed propagation. Aquac. Sci., 46, 97–100. https://doi.org/10.11233/aquaculturesci1953.46.97

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

Money, D., Gardner, K., Migicovsky, Z., Schwaninger, H., Zhong, G.Y., Myles, S., 2015. LinkImpute: Fast and accurate genotype imputation for nonmodel organisms. G3 Genes, Genomes, Genet., 5, 2383–2390. https://doi.org/10.1534/g3.115.021667

Money, D., Migicovsky, Z., Gardner, K., Myles, S., 2017. LinkImputeR: User-guided genotype calling and imputation for non-model organisms. BMC Genomics, 18, 523. https://doi.org/10.1186/s12864-017-3873-5

Motte, C., Rios, A., Lefebvre, T., Do, H., Henry, M., Jintasataporn, O., 2019. Replacing fish meal with defatted insect meal (*Yellow mealworm tenebrio molitor*) improves the growth and immunity of pacific white shrimp (*litopenaeus vannamei*). Animals, 9, 258. https://doi.org/10.3390/ani9050258

Mulder, H.A., Bijma, P., 2005. Effects of genotype x environment interaction on genetic gain in breeding programs. J. Anim. Sci., 83(1), 49–61. https://doi.org/10.2527/2005.83149x

Murashita, K., Matsunari, H., Fukada, H., Suzuki, N., Furuita, H., Oku, H., Rønnestad, I., Yoshinaga, H., Yamamoto, T., 2019. Effect of a plant-based low-fishmeal diet on digestive physiology in yellowtail *Seriola quinqueradiata*. Aquaculture, 506(15), 168-180. https://doi.org/10.1016/j.aquaculture.2019.03.040

Neira, R., 2010. Breeding in aquaculture species : genetic improvement programs in developing countries, in: The 9th World Congress on Genetics Applied to Livestock Production. p. 8.

Noguchi, T., Takada, Y., Fujiwara, S., 1979. Degradation of uric acid to urea and glyoxylate in peroxisomes. J. Biol. Chem., 254(12), 5272-5275. https://doi.org/10.1016/s0021-9258(18)50590-1

Norambuena, F., Lewis, M., Hamid, N.K.A., Hermon, K., Donald, J.A., Turchini, G.M., 2013. Fish oil replacement in current aquaculture feed: Is cholesterol a hidden treasure for fish nutrition? PLoS One, 8(12), e81705. https://doi.org/10.1371/journal.pone.0081705

Ødegård, J., Baranski, M., Gjerde, B., Gjedrem, T., 2011. Methodology for genetic evaluation of disease resistance in aquaculture species: challenges and future prospects. Aquac. Res., 42, 103–114. https://doi.org/10.1111/j.1365-2109.2010.02669.x

Odegård, J., Moen, T., Santi, N., Korsvoll, S.A., Kjøglum, S., Meuwissen, T.H.E., 2014. Genomic prediction in an admixed population of Atlantic salmon (*Salmo salar*). Front. Genet., 5, 402. https://doi.org/10.3389/fgene.2014.00402

Ogawa, K., 2016. Heterobothriosis of cultured Japanese pufferfish *Takifugu rubripes*. Fish Pathol., 51(2), 39–43.

Ogawa, K., 2002. Impacts of diclidophorid monogenean infections on fisheries in Japan, in: International Journal for Parasitology. Int J Parasitol, 32(3), 373–380. https://doi.org/10.1016/S0020-7519(01)00338-1

Ogawa, K., Inouye, K., 1997. Heterobothrium Infection of Cultured Tiger Puffer, *Takifugu rubripes*. A Field Observation. Fish Pathol., 32,

15–20. https://doi.org/10.3147/jsfp.32.15

Oldenbroek, K., van der Waaij, L., 2015. Textbook Animal Breeding and Genetics for BSc students. Centre for Genetic Resources and Animal Breeding and Genomics Group, Wageningen University and Research Centre

Øverland, M., Karlsson, A., Mydland, L.T., Romarheim, O.H., Skrede, A., 2013. Evaluation of Candida utilis, Kluyveromyces marxianus and *Saccharomyces cerevisiae* yeasts as protein sources in diets for Atlantic salmon (*Salmo salar*). Aquaculture, 402–403, 1-7. https://doi.org/10.1016/j.aquaculture.2013.03.016

Overturf, K., Barrows, F.T., Hardy, R.W., 2013. Effect and interaction of rainbow trout strain (Oncorhynchus mykiss) and diet type on growth and nutrient retention. Aquac. Res., 44(4), 604-611. https://doi.org/10.1111/j.1365-2109.2011.03065.x

Ozório, R.O.A., Portz, L., Borghesi, R., Cyrino, J.E.P., 2012. Effects of dietary yeast (*saccharomyces cerevisia*) supplementation in practical diets of tilapia (*oreochromis niloticus*). Animals, 2(1), 16-24. https://doi.org/10.3390/ani2010016

Palaiokostas, C., Cariou, S., Bestin, A., Bruant, J.S., Haffray, P., Morin, T., Cabon, J., Allal, F., Vandeputte, M., Houston, R.D., 2018. Genome-wide association and genomic prediction of resistance to viral nervous necrosis in European sea bass (*Dicentrarchus labrax*) using RAD sequencing. Genet. Sel. Evol., 50, 30. https://doi.org/10.1186/s12711-018-0401-2

Palaiokostas, C., Ferraresso, S., Franch, R., Houston, R.D., Bargelloni, L., 2016. Genetics of resistance to photobacteriosis in gilthead sea bream (*Sparus aurata*) using 2b-RAD sequencing. G3 (Bethesda), 6, 3693–3700. https://doi.org/10.1534/g3.116.035220

Pandey, S., Takahama, M., Gruenbaum, A., Zewde, M., Cheronis, K., Chevrier, N., 2020. A whole-tissue RNA-seq toolkit for organism-wide studies of gene expression with PME-seq. Nat. Protoc., 15, 1459–1483. https://doi.org/10.1038/s41596-019-0291-y

Park, T., Casella, G., 2008. The Bayesian Lasso. J. Am. Stat. Assoc., 103, 681–686. https://doi.org/10.1198/016214508000000337

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., Duchesnay, E., 2011. Scikit-learn: Machine learning in Python. J. Mach. Learn. Res., 12, 2825–2830.

Perera, W.M.K., Carter, C.G., Houlihan, D.F., 1995. Feed consumption, growth and growth efficiency of rainbow trout (*Oncorhynchus mykiss* (Walbaum)) fed on diets containing a bacterial single-cell protein . Br. J. Nutr., 73(4), 591 - 603. https://doi.org/10.1079/bjn19950061

Pérez-Enciso, M., Zingaretti, L.M., 2019. A guide for using deep learning for complex trait genomic prediction. Genes (Basel), 10(7), 553. https://doi.org/10.3390/genes10070553

Pérez, P., de los Campos, G., 2014. Genome-wide regression and prediction with the BGLR statistical package. Genetics, 198, 483–495. https://doi.org/10.1534/genetics.114.164442

Peterson, B.C., Booth, N.J., Manning, B.B., 2012. Replacement of fish meal in juvenile channel catfish, *Ictalurus punctatus*, diets using a yeast-derived protein source: The effects on weight gain, food conversion ratio, body composition and survival of catfish challenged with *Edwardsiella ictaluri*. Aquac. Nutr., 18(2), 132-137. https://doi.org/10.1111/j.1365-2095.2011.00878.x

Phipson, B., Lee, S., Majewski, I.J., Alexander, W.S., Smyth, G.K., 2016. Robust hyperparameter estimation protects against hypervariable genes and improves power to detect differential expression. Ann. Appl. Stat., 10(2), 946-963. https://doi.org/10.1214/16-AOAS920

Pierce, L.R., Palti, Y., Silverstein, J.T., Barrows, F.T., Hallerman, E.M., Parsons, J.E., 2008. Family growth response to fishmeal and plant-based diets shows genotype × diet interaction in rainbow trout (*Oncorhynchus mykiss*). Aquaculture 278(1-4), 37-42. https://doi.org/10.1016/j.aquaculture.2008.03.017

Poplin, R., Ruano-Rubio, V., DePristo, M.A., Fennell, T.J., Carneiro, M.O., Auwera, G.A. Van der, Kling, D.E., Gauthier, L.D., Levy-Moonshine, A., Roazen, D., Shakir, K., Thibault, J., Chandran, S., Whelan, C., Lek, M., Gabriel, S., Daly, M.J., Neale, B., MacArthur, D.G., Banks, E., 2017. Scaling accurate genetic variant discovery to tens of thousands of samples. bioRxiv, 201178. https://doi.org/10.1101/201178

Price, A.L., Zaitlen, N.A., Reich, D., Patterson, N., 2010. New approaches to population stratification in genome-wide association studies. Nat. Rev. Genet., 11, 459–63. 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., Sham, P.C., 2007. PLINK: A tool set for whole-genome association and population-based linkage analyses. Am. J. Hum. Genet., 81, 559–575. https://doi.org/10.1086/519795

Rauw, W.M., Gomez-Raya, L., 2015. Genotype by environment interaction and breeding for robustness in livestock. Front. Genet., 6. https://doi.org/10.3389/fgene.2015.00310

Reich, D., Price, A.L., Patterson, N., 2008. Principal component analysis of genetic data. Nat. Genet., 40, 491–492 https://doi.org/10.1038/ng0508-491

Rhodes, M.A., Zhou, Y., Davis, D.A., 2015. Use of dried fermented biomass as a fish meal replacement in practical diets of Florida pompano, *Trachinotus carolinus*. J. Appl. Aquac., 27(1), 29-39. https://doi.org/10.1080/10454438.2014.959834

Rimoldi, S., Antonini, M., Gasco, L., Moroni, F., Terova, G., 2021. Intestinal microbial communities of rainbow trout (*Oncorhynchus mykiss*) may be improved by feeding a *Hermetia illucens* meal/low-fishmeal diet. Fish Physiol. Biochem., 47, 365–380. https://doi.org/10.1007/s10695-020-00918-1

Rimoldi, S., Torrecillas, S., Montero, D., Gini, E., Makol, A., Victoria Valdenegro, V., Izquierdo, M., Terova, G., 2020. Assessment of

dietary supplementation with galactomannan oligosaccharides and phytogenics on gut microbiota of European sea bass (*Dicentrarchus Labrax*) fed low fishmeal and fish oil based diet. PLoS One, 15, e0231494. https://doi.org/10.1371/journal.pone.0231494

Ritchie, M.E., Phipson, B., Wu, D., Hu, Y., Law, C.W., Shi, W., Smyth, G.K., 2015. Limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res., 43, e47. https://doi.org/10.1093/nar/gkv007

Ritz, C., Spiess, A.N., 2008. qpcR: An R package for sigmoidal model selection in quantitative real-time polymerase chain reaction analysis. Bioinformatics 24(13), 1549–1551. https://doi.org/10.1093/bioinformatics/btn227

Robinson, M.D., Oshlack, A., 2010. A scaling normalization method for differential expression analysis of RNA-seq data. Genome Biol., 11, R25. https://doi.org/10.1186/gb-2010-11-3-r25

Robledo, D., Matika, O., Hamilton, A., Houston, R.D., 2018a. Genome-wide association and genomic selection for resistance to amoebic gill disease in Atlantic salmon. G3 Genes, Genomes, Genet., 8, 1195–1203. https://doi.org/10.1534/g3.118.200075

Robledo, D., Palaiokostas, C., Bargelloni, L., Martínez, P., Houston, R., 2018b. Applications of genotyping by sequencing in aquaculture breeding and genetics. Rev. Aquac. 10(3), 670-682. https://doi.org/10.1111/raq.12193

Romarheim, O.H., Øverland, M., Mydland, L.T., Skrede, A., Landsverk, T., 2011. Bacteria grown on natural gas prevent soybean meal-induced enteritis in atlantic salmon. J. Nutr. 141(1), 124–130. https://doi.org/10.3945/jn.110.128900

Roques, S., Deborde, C., Richard, N., Sergent, L., Kurz, F., Skiba-Cassy, S., Fauconneau, B., Moing, A., 2018. Characterizing alternative feeds for rainbow trout (*O. mykiss*) by ^{1}H NMR metabolomics. Metabolomics, 14, 155. https://doi.org/10.1007/s11306-018-1454-5

Rumsey, G.L., Kinsella, J.E., Shetty, K.J., Hughes, S.G., 1991. Effect of high dietary concentrations of brewer's dried yeast on growth performance and liver uricase in rainbow trout (*Oncorhynchus mykiss*). Anim. Feed Sci. Technol., 33, 177–183. https://doi.org/10.1016/0377-8401(91)90058-Z

Rye, M., Gjerde, B., Gjedrem, T., 2010. Genetic improvement programs for aquaculture species in developed countries, in: The 9th World Congress on Genetics Applied to Livestock Production. p. 8.

Sae-Lim, P., Kause, A., Janhunen, M., Vehviläinen, H., Koskinen, H., Gjerde, B., Lillehammer, M., Mulder, H.A., 2015. Genetic (co)variance of rainbow trout (*Oncorhynchus mykiss*) body weight and its uniformity across production environments. Genet. Sel. Evol., 47, 46. https://doi.org/10.1186/s12711-015-0122-8

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. J. Anim. Sci., 91(12), 5572–5581. https://doi.org/10.2527/jas.2012-5949

Sánchez-Muniz, F.J., de La Higuera, M., Varela, G., 1982. Alterations of erythrocytes of the rainbow trout (*Salmo gairdneri*) by the use of hansenula anomala yeast as sole protein source. Comp. Biochem. Physiol. -- Part A Physiol, 72(4), 693-696. https://doi.org/10.1016/0300-9629(82)90150-5

Sato, M., Hosoya, S., Yoshikawa, S., Ohki, S., Kobayashi, Y., Itou, T., Kikuchi, K., 2019. A highly flexible and repeatable genotyping method for aquaculture studies based on target amplicon sequencing using next-generation sequencing technology. Sci. Rep., 9, 6904. https://doi.org/10.1038/s41598-019-43336-x

Satoh, S., Hernández, A., Tokoro, T., Morishita, Y., Kiron, V., Watanabe, T., 2003. Comparison of phosphorus retention efficiency between rainbow trout (*Oncorhynchus mykiss*) fed a commercial diet and a low fish meal based diet. Aquaculture, 224, 271–282. https://doi.org/10.1016/S0044-8486(03)00217-5

Seong, T., Matsutani, H., Haga, Y., Kitagima, R., Satoh, S., 2019. First step of non-fish meal, non-fish oil diet development for red seabream, (*Pagrus major*), with plant protein sources and microalgae Schizochytrium sp. Aquac. Res., 50, 2460–2468. https://doi.org/10.1111/are.14199

Serradell, A., Torrecillas, S., Makol, A., Valdenegro, V., Fernández-Montero, A., Acosta, F., Izquierdo, M.S., Montero, D., 2020. Prebiotics and phytogenics functional additives in low fish meal and fish oil based diets for European sea bass (*Dicentrarchus labrax*): Effects on stress and immune responses. Fish Shellfish Immunol., 100, 219–229. https://doi.org/10.1016/j.fsi.2020.03.016

Shirakashi, S., Nakatsuka, S., Udagawa, A., Ogawa, K., 2010. Oncomiracidial behavior of *Heterobothrium okamotoi* (Monogenea: Diclidophoridae). Fish Pathol. 45, 51–57. https://doi.org/10.3147/jsfp.45.51

Sibert, J.R., Nielsen, J.L., 2001. Electronic tagging and tracking in marine fisheries : proceedings of the symposium on tagging and tracking marine fish with electronic devices, February 7-11, 2000, East-West Center, University of Hawaii. Springer Netherlands.

Skoglund, P., Götherström, A., Jakobsson, M., 2011. Estimation of population divergence times from non-overlapping genomic sequences: Examples from dogs and wolves. Mol. Biol. Evol. 28(4), 1505–1517. https://doi.org/10.1093/molbev/msq342

Smith, H.F., 1936. A discriminant function for plant selection. Ann. Eugen., 7, 240–250. https://doi.org/10.1111/j.1469-1809.1936.tb02143.x

Smyth, G.K., 2004. Linear models and empirical bayes methods for assessing differential expression in microarray experiments. Stat. Appl. Genet. Mol. Biol., 3(1). https://doi.org/10.2202/1544-6115.1027

Smyth, G.K., Ritchie, M.E., Law, C.W., Alhamdoosh, M., Su, S., Dong, X., Tian, L., 2018. RNA-seq analysis is easy as 1-2-3 with limma, Glimma and edgeR. F1000Research, 5, ISCB Comm J-1408. https://doi.org/10.12688/f1000research.9005.3

Solberg, T.R., Sonesson, A.K., Woolliams, J.A., Meuwissen, T.H.E., 2008. Genomic selection using different marker types and densities. J. Anim. Sci., 86, 2447–2454. https://doi.org/10.2527/jas.2007-0010

Stark, R., Grzelak, M., Hadfield, J., 2019. RNA sequencing: the teenage years. Nat. Rev. Genet., 20(11), 631-656. https://doi.org/10.1038/s41576-019-0150-2

Strucken, E.M., Laurenson, Y.C.S.M., Brockmann, G.A., 2015. Go with the flow-biology and genetics of the lactation cycle. Front. Genet., 6. https://doi.org/10.3389/fgene.2015.00118

Sun, H., Zhang, A. hua, Song, Q., Fang, H., Liu, X. yuan, Su, J., Yang, L., Yu, M. die, Wang, X. jun, 2018. Functional metabolomics discover pentose and glucuronate interconversion pathways as promising targets for Yang Huang syndrome treatment with Yinchenhao Tang. RSC Adv. 8. https://doi.org/10.1039/C8RA06553E

Takeuchi, Y., Yoshizaki, G., Takeuchi, T., 2004. Surrogate broodstock produces salmonids. Nature, 430, 629–630. https://doi.org/10.1038/430629a

Takii, K., Akiyama, S., Maoka, T., Hidaka, K., Otaka, K., Seoka, M., Kumai, H., 2004. Evaluation of dietary yeast autolysates for red sea bream, *Pagrus major*. Eval. Diet., 52, 387–394. https://doi.org/10.11233/aquaculturesci1953.52.387

Tan, B., Mai, K., Zheng, S., Zhou, Q., Liu, L., Yu, Y., 2005. Replacement of fish meal by meat and bone meal in practical diets for the white shrimp *Litopenaeus vannamai* (Boone). Aquac. Res., 36, 439–444. https://doi.org/10.1111/j.1365-2109.2005.01223.x

Teressa, T., Semahegn, Z., Bejiga, T., 2021. Multi Environments and Genetic-Environmental Interaction (GxE) in Plant Breeding and its Challenges: A Review Article. Int. J. Res. Stud. Agric. Sci., 7(4), 11-18. https://doi.org/10.20431/2454-6224.0704002

Thorpe, J.L., Doitsidou, M., Ho, S.Y., Raz, E., Farber, S.A., 2004. Germ cell migration in zebrafish is dependent on HMGCoA reductase activity and prenylation. Dev. Cell, 6(2), 295-302. https://doi.org/10.1016/S1534-5807(04)00032-2

Tiezzi, F., Maltecca, C., 2015. Accounting for trait architecture in genomic predictions of US Holstein cattle using a weighted realized relationship matrix. Genet. Sel. Evol., 47, 24. https://doi.org/10.1186/s12711-015-0100-1

Togashi, K., Lin, C.Y., Yamazaki, T., 2011. The efficiency of genome-wide selection for genetic improvement of net merit. J. Anim. Sci., 89, 2972–2980. https://doi.org/10.2527/jas.2009-2606

Tsai, H.Y., Hamilton, A., Tinch, A.E., Guy, D.R., Bron, J.E., Taggart, J.B., Gharbi, K., Stear, M., Matika, O., Pong-Wong, R., Bishop, S.C., Houston, R.D., 2016. Genomic prediction of host resistance to sea lice in farmed Atlantic salmon populations. Genet. Sel. Evol., 48, 47. https://doi.org/10.1186/s12711-016-0226-9

Tsai, H.Y., Hamilton, A., Tinch, A.E., Guy, D.R., Gharbi, K., Stear, M.J., Matika, O., Bishop, S.C., Houston, R.D., 2015. Genome wide association and genomic prediction for growth traits in juvenile farmed Atlantic salmon using a high density SNP array. BMC Genomics, 16, 969. https://doi.org/10.1186/s12864-015-2117-9

Tsairidou, S., Hamilton, A., Robledo, D., Bron, J.E., Houston, R.D., 2020. Optimizing low-cost genotyping and imputation strategies for genomic selection in atlantic salmon. G3 Genes, Genomes, Genet., 10, 581–590. https://doi.org/10.1534/g3.119.400800

Uffelmann, E., Huang, Q.Q., Munung, N.S., de Vries, J., Okada, Y., Martin, A.R., Martin, H.C., Lappalainen, T., Posthuma, D., 2021. Genome-wide association studies. Nat. Rev. Methods Prim.,1, 1–21. https://doi.org/10.1038/s43586-021-00056-9

Ukawa, M., Takii, K., Nakamura, M., Kumai, H., 1996. Utilization of some protein sources in single moist pellets for tiger puffer. Aquac. Sci., 44, 511–516. https://doi.org/10.11233/aquaculturesci1953.44.511

Vallejo, R.L., Leeds, T.D., Gao, G., Parsons, J.E., Martin, K.E., Evenhuis, J.P., Fragomeni, B.O., Wiens, G.D., Palti, Y., 2017. Genomic selection models double the accuracy of predicted breeding values for bacterial cold water disease resistance compared to a traditional pedigree-based model in rainbow trout aquaculture. Genet. Sel. Evol., 49, 1–13. https://doi.org/10.1186/s12711-017-0293-6

Van Der Maaten, L.J.P., Hinton, G.E., 2008. Visualizing high-dimensional data using t-sne. J. Mach. Learn. Res., 9, 2579–2605. https://doi.org/10.1007/s10479-011-0841-3

Vapnik, V.N., 1995. The Nature of Statistical Learning Theory. AT&T Bell Laboratories, Holmdel, USA. https://doi.org/10.1007/978-1-4757-2440-0

Veeranagouda, Y., Remaury, A., Guillemot, J.C., Didier, M., 2019. RNA Fragmentation and Sequencing (RF-Seq): cost-effective, time-efficient, and high-throughput 3′ mRNA sequencing library construction in a single tube. Curr. Protoc. Mol. Biol. 129, e109. https://doi.org/10.1002/cpmb.109

Vidakovic, A., Huyben, D., Sundh, H., Nyman, A., Vielma, J., Passoth, V., Kiessling, A., Lundh, T., 2020. Growth performance, nutrient digestibility and intestinal morphology of rainbow trout (*Oncorhynchus mykiss*) fed graded levels of the yeasts *Saccharomyces cerevisiae* and *Wickerhamomyces anomalus*. Aquac. Nutr., 26(2), 275-286. https://doi.org/10.1111/anu.12988

Visscher, P.M., Hill, W.G., Wray, N.R., 2008. Heritability in the genomics era - Concepts and misconceptions. Nat. Rev. Genet., 9, 255–266. https://doi.org/10.1038/nrg2322

Wakchaure, R., Ganguly, S., 2015. Marker Assisted Selection (MAS) in animal breeding: A Review. J. Drug Metab. Toxicol., 6, 5. https://doi.org/10.4172/2157-7609.1000e127

Walsh, B., Lynch, M., 2000. 8 Family-Based Selection.

Wang, C., Chuprom, J., Wang, Y., Fu, L., 2020. Beneficial bacteria for aquaculture: nutrition, bacteriostasis and immunoregulation. J. Appl. Microbiol., 128, 28–40. https://doi.org/10.1111/jam.14383

Wang, Q., Yu, Y., Zhang, Q., Zhang, X., Huang, H., Xiang, J., Li, F., 2019. Evaluation on the genomic selection in *Litopenaeus vannamei* for the resistance against *Vibrio parahaemolyticus*. Aquaculture, 505, 212–216. https://doi.org/10.1016/j.aquaculture.2019.02.055

Widmer, C., Rätsch, G., 2011. Multitask learning in computational biology, in: The 2011 International Conference on Unsupervised and Transfer Learning Workshop. 207–216.

Wu, T., Hu, E., Xu, S., Chen, M., Guo, P., Dai, Z., Feng, T., Zhou, L., Tang, W., Zhan, L., Fu, X., Liu, S., Bo, X., Yu, G., 2021. clusterProfiler 4.0: A universal enrichment tool for interpreting omics data. Innov., 2, 100141. https://doi.org/10.1016/J.XINN.2021.100141

Xie, S. wei, Liu, Y. jian, Zeng, S., Niu, J., Tian, L. xia, 2016. Partial replacement of fish-meal by soy protein concentrate and soybean meal based protein blend for juvenile Pacific white shrimp, *Litopenaeus vannamei*. Aquaculture, 464, 296–302. https://doi.org/10.1016/j.aquaculture.2016.07.002

Xie, S. wei, Tian, L. xia, Jin, Y., Yang, H. jun, Liang, G. ying, Liu, Y. jian, 2014. Effect of glycine supplementation on growth performance, body composition and salinity stress of juvenile Pacific white shrimp, *Litopenaeus vannamei* fed low fishmeal diet. Aquaculture, 418–419, 159–164. https://doi.org/10.1016/j.aquaculture.2013.10.023

Xu, Y., Liu, X., Fu, J., Wang, H., Wang, J., Huang, C., Prasanna, B.M., Olsen, M.S., Wang, G., Zhang, A., 2020. Enhancing genetic gain through genomic sselection: from livestock to plants. Plant Commun., 1(1), 13. https://doi.org/10.1016/j.xplc.2019.100005

Yamamoto, T., Murashita, K., Matsunari, H., Oku, H., Furuita, H., Okamoto, H., Amano, S., Suzuki, N., 2016. Amago salmon *Oncorhynchus masou ishikawae* juveniles selectively bred for growth on a low fishmeal diet exhibit a good response to the low fishmeal diet due largely to an increased feed intake with a particular preference for the diet. Aquaculture, 465, 380–386. https://doi.org/10.1016/j.aquaculture.2016.09.030

Yamamoto, T., Murashita, K., Matsunari, H., Oku, H., Furuita, H., Okamoto, H., Amano, S., Suzuki, N., 2015. Selectively bred juvenile F2 amago salmon *Oncorhynchus masou ishikawae* fed a low fishmeal diet exhibit growth comparable to unselected juveniles fed a fishmeal-based diet. Fish. Sci., 81, 83–93. https://doi.org/10.1007/s12562-014-0817-8

Yamamoto, T., Okamoto, H., Furuita, H., Murashita, K., Matsunari, H., Iwashita, Y., Amano, S., Suzuki, N., 2014. Growth performance and physiological condition of F1 amago salmon *Oncorhynchus masou ishikawae* juveniles obtained from broodstock with selective breeding for growth on a low fish-meal diet. Fish. Sci. 80. https://doi.org/10.1007/s12562-014-0707-0

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

Yoshikawa, S., Chuda, H., Hamasaki, M., Kadomura, K., Yamada, T., Kikuchi, K., Hosoya, S., 2020. Precocious maturation in male tiger pufferfish *Takifugu rubripes*: genetics and endocrinology. Fish. Sci., 86, 339–351. https://doi.org/10.1007/s12562-019-01390-4

Yoshikawa, S., Hamasaki, M., Kadomura, K., Yamada, T., Chuda, H., Kikuchi, K., Hosoya, S., 2021. Genetic Dissection of a Precocious Phenotype in Male Tiger Pufferfish (*Takifugu rubripes*) using Genotyping by Random Amplicon Sequencing, Direct (GRAS-Di). Mar. Biotechnol. 23, 177–188. https://doi.org/10.1007/s10126-020-10013-4

Yoshitomi, B., Aoki, M., Oshima, S. ichirou, 2007. Effect of total replacement of dietary fish meal by low fluoride krill (*Euphausia superba*) meal on growth performance of rainbow trout (*Oncorhynchus mykiss*) in fresh water. Aquaculture, 266, 219–225. https://doi.org/10.1016/j.aquaculture.2006.12.043

Yoshizaki, G., Yazawa, R., 2019. Application of surrogate broodstock technology in aquaculture. Fish. Sci., 85, 429-437. https://doi.org/10.1007/s12562-019-01299-y

Yu, G., Wang, L.G., Han, Y., He, Q.Y., 2012. ClusterProfiler: An R package for comparing biological themes among gene clusters. Omi. A J. Integr. Biol., 16, 284–287. https://doi.org/10.1089/omi.2011.0118

Zenger, K.R., Khatkar, M.S., Jones, D.B., Khalilisamani, N., Jerry, D.R., Raadsma, H.W., 2019. Genomic selection in aquaculture: Application, limitations and opportunities with special reference to marine shrimp and pearl oysters. Front. Genet., 9. https://doi.org/10.3389/fgene.2018.00693

Zhang, X., Belsky, J., 2020. Three phases of Gene × Environment interaction research: Theoretical assumptions underlying gene selection. Dev. Psychopathol., First View, 1-12. https://doi.org/10.1017/S0954579420000966

Zhao, L., Wang, W., Huang, X., Guo, T., Wen, W., Feng, L., Wei, L., 2017. The effect of replacement of fish meal by yeast extract on the digestibility, growth and muscle composition of the shrimp *Litopenaeus vannamei*. Aquac. Res., 48, 311–320. https://doi.org/10.1111/are.12883

Figure and Table

Table 1-1

Predictive ability (mean ± standard error) on *Heterobothrium okamotoi* count (HC) and standard length (SL) under 12 models: GBLUP, Bayes A, Bayes B, Bayes C, Bayes LASSO, Bayes reproducing kernel Hilbert space (Bayes RKHS), support vector machine with a linear kernel (SVR-linear), SVR with a poly kernel (SVR-poly), SVR with a radial basis function kernel (SVR-rbf), feedforward neural networks (FNN), and multi-task feedforward neural networks (multi-task FNN). The top three models for HC and SL are highlighted with bold font

Model	**HC**	**SL**
GBLUP	0.307 ± 0.018	0.463 ± 0.018
Bayes A	0.312 ± 0.018	0.461 ± 0.018
Bayes B	0.306 ± 0.018	0.460 ± 0.018
Bayes C	0.307 ± 0.018	0.460 ± 0.018
Bayes LASSO	0.303 ± 0.018	**0.464 ± 0.018**
Bayes ridge	0.304 ± 0.018	0.460 ± 0.018
Bayes RKHS	**0.325 ± 0.019**	0.463 ± 0.018
SVR-linear	0.322 ± 0.017	0.410 ± 0.019
SVR-poly	0.248 ± 0.019	**0.481 ± 0.018**
SVR-rbf	0.249 ± 0.019	**0.475 ± 0.018**
FNN	**0.330 ± 0.018**	0.405 ± 0.017
Multi-task FNN	**0.344 ± 0.019**	0.340 ± 0.022

Figure 1-1

Histograms with the estimated density of phenotypes: (**a**) *Heterobothrium okamotoi* count (HC), (**b**) transformed HC, and (**c**) standard length (SL).

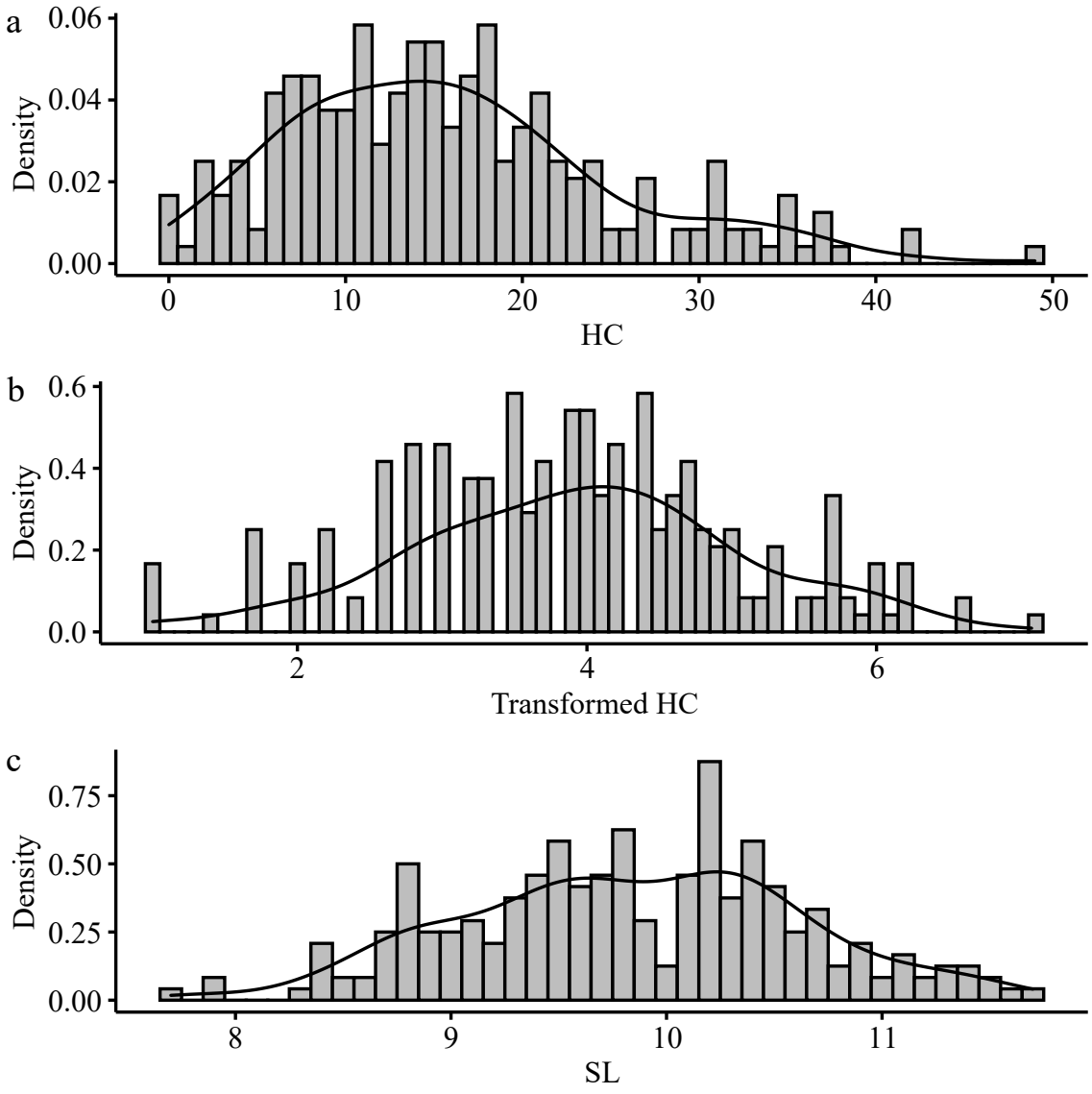

Figure 1-2

Population structure detected by *t*-SNE analysis based on the genomic SNP data of each individual (filled circle). Filled colors represent *Heterobothrium okamotoi* count of each individual based on the color bar (right panel).

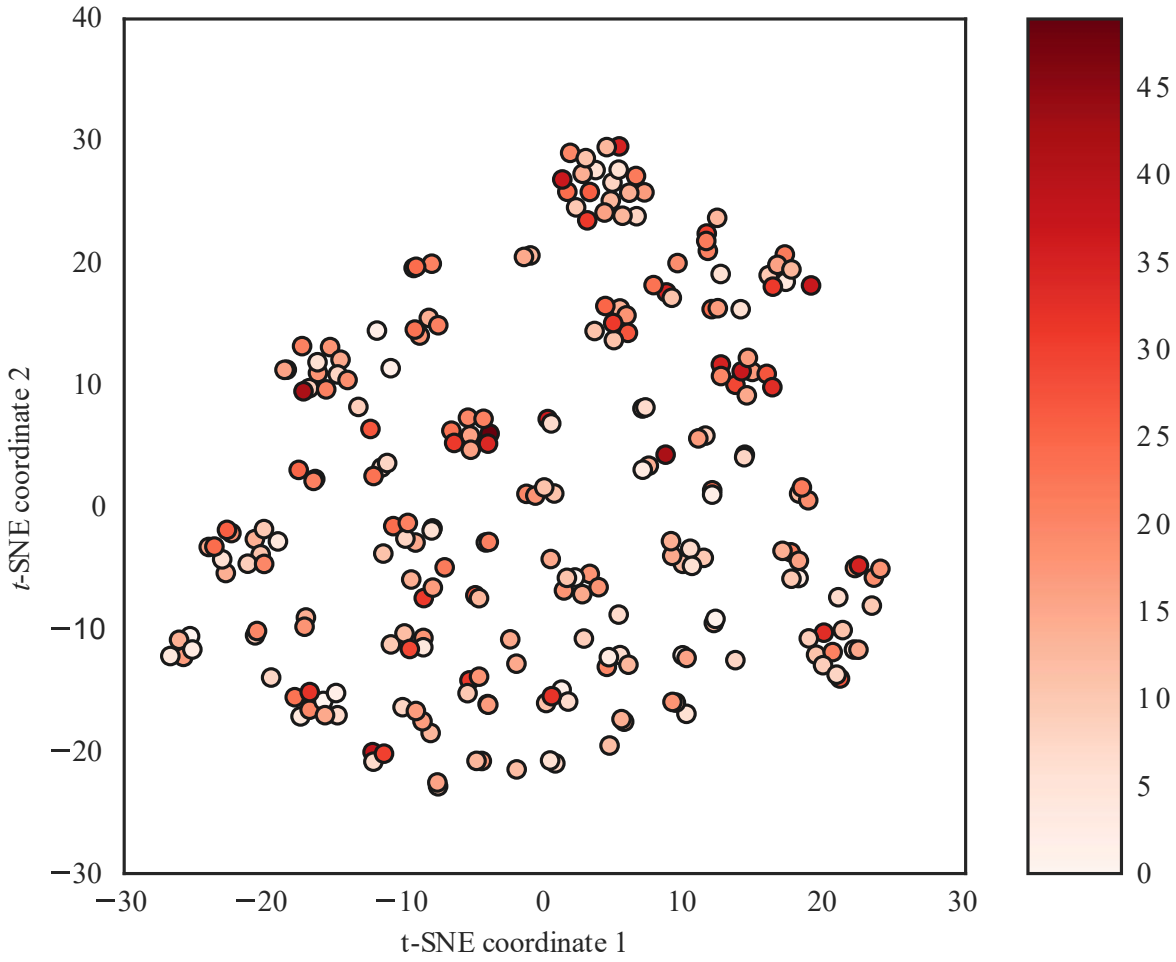

Figure 1-3

Manhattan plots from genome-wide association study: (**a**) transformed *Heterobothrium okamotoi* count (HC), and (**b**) standard length (SL). Adjacent chromosomes are distinguished by different colors. The X-axis is the physical order of the SNP markers across the 22 chromosomes of *Takifugu rubripes*. The Y-axis represents the negative logarithm of *p*-values (base: 10) for the target trait. Red dashed lines are Bonferroni-corrected significance thresholds of 5.128.

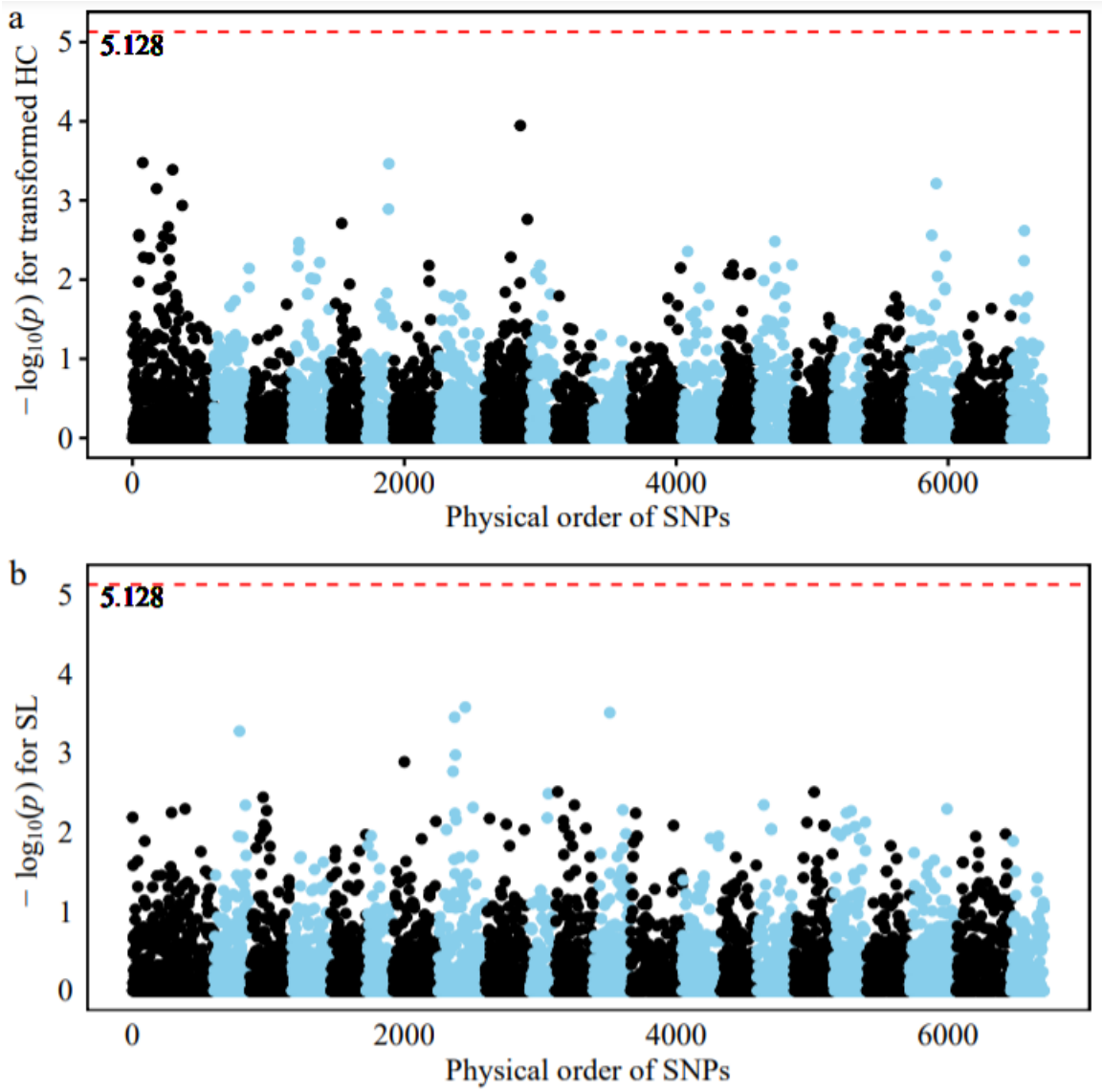

Figure 1-4

The diagrams of the simulation study. (**a**) The initiation of the breeding program shared among all scenarios. The founder population (n = 10,000) was constructed, and 20 sires and 20 dams were randomly sampled to produce 8000 progenies. Then, 2000 fish were randomly sampled from the progeny pool as the broodstock population (F_0). (**b**) The workflow of recurrent selection schemes. Parents (20 sires and 20 dams) were selected from F_0 according to the scenario-specific selection criteria and 8000 progenies were generated. The selection scenarios were: RAND, random selection; GS_{HC}, selection on *Heterobothrium okamotoi* counts (HC); GS_{SL}, selection on standard length (SL); $S1_{SHI}$ and $S2_{SHI}$, selection based on genomic Smith-Hazel index (SHI); S_{DGI}, selection based on the desired gains index (DGI). $S1_{SHI}$ has the same economic weights for both traits, and $S2_{SHI}$ uses the similar vector of economic weights as the vector of desired gains in S_{DGI}. Then, random sampling was applied to select 2000 progenies as the broodstock population for the next generation. A total of 10 generations (F_1 to F_{10}) of this process were replicated 50 times.

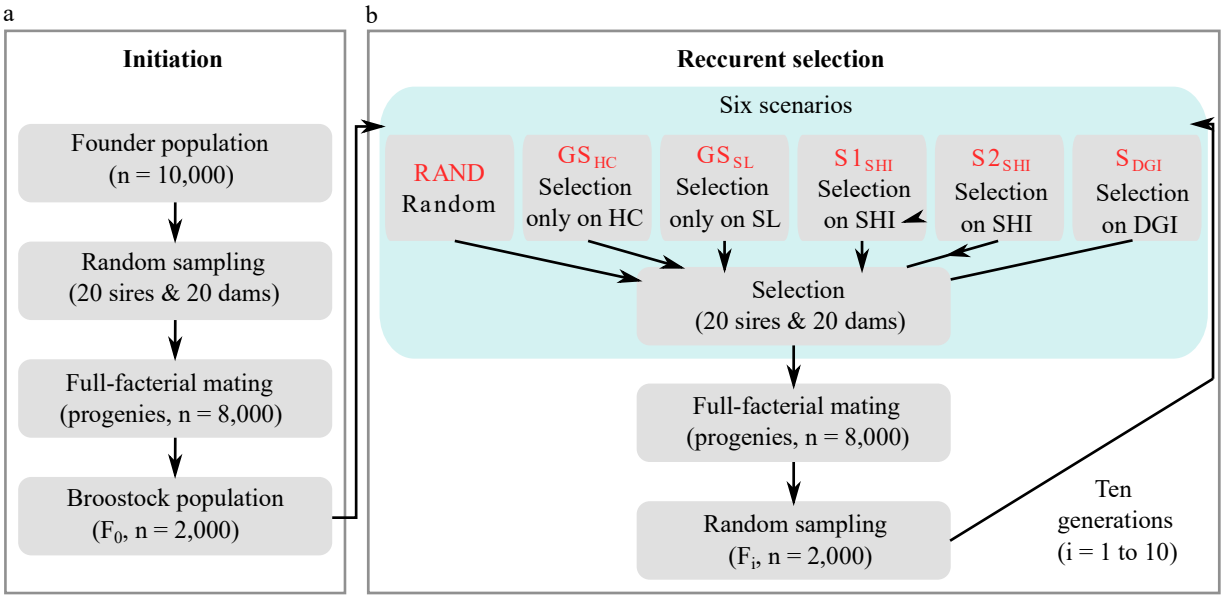

Figure 1-5

Genetic trends of average true breeding value (TBV) for *Heterobothrium okamotoi* count (HC, red lines) and standard length (SL, blue lines) of broodstock population in each generation (F_0 to F_{10}) among five different simulation scenarios with 50 replicates. (**a**) random mating (RAND), (**b**) GS on HC only (GS_{HC}), (**c**) GS on SL only (GS_{SL}), (**d**) Smith-Hazel index with the same economic weights $S1_{SHI}$), (**e**) Smith-Hazel index with the different economic weights ($S2_{SHI}$), and (**f**) desired gains index (S_{DGI}).

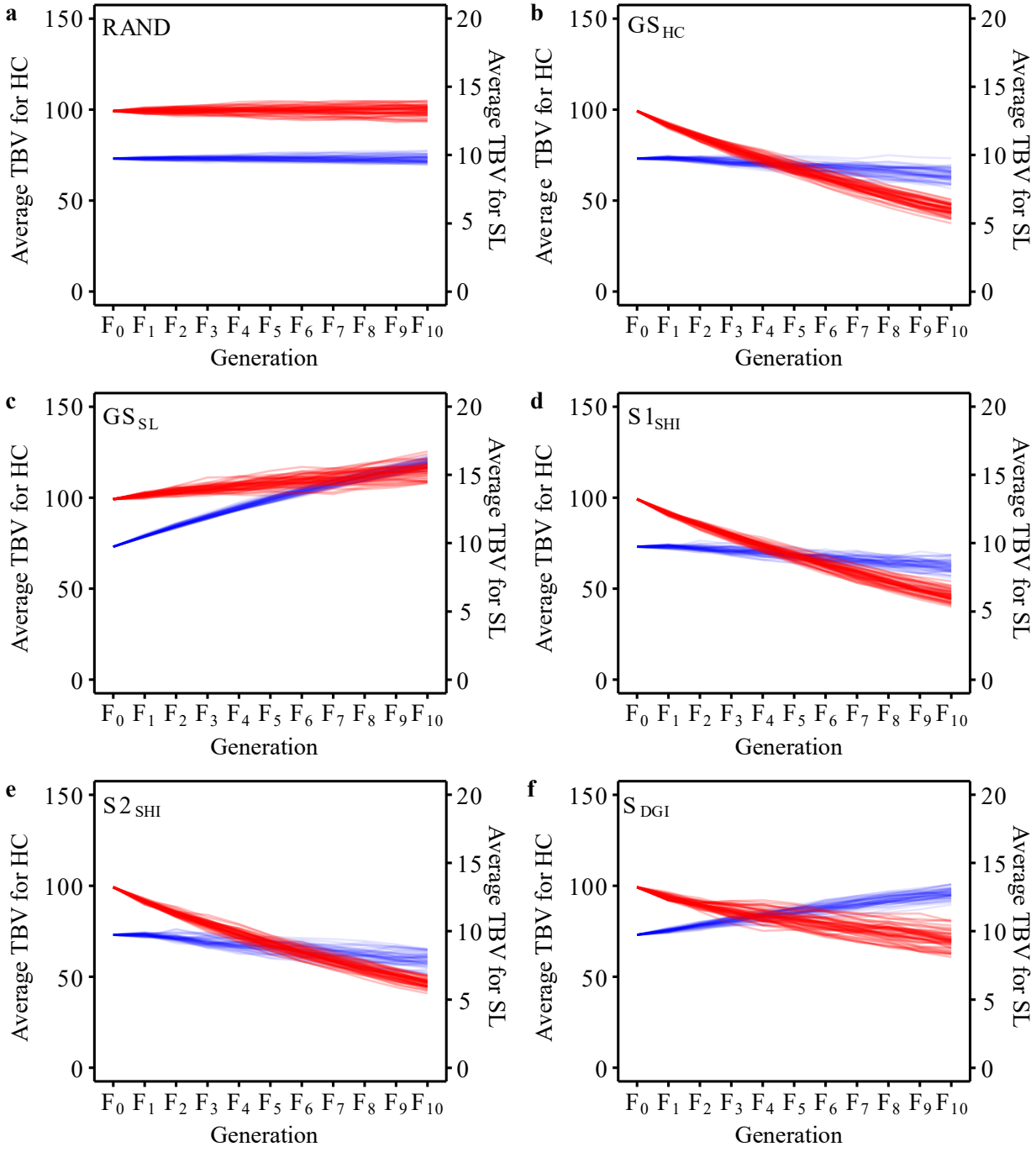

Table 2-1

Predicitive abilies (mean, S.D. and S.E.M., 50 replicates) of genomic prediction for *Heterobothrium okamotoi* count (HC) and standard length using GBLUP model with different number of SNPs (50, 100, 500, 1,000, 2,500, 5,000, 7,500 and 10,000 SNPs)

SNP number	HC			SL		
	Mean	**S.D.**	**S.E.M.**	**Mean**	**S.D.**	**S.E.M.**
50	0.168	0.021	0.003	0.258	0.037	0.005
100	0.194	0.019	0.003	0.314	0.034	0.005
500	0.232	0.013	0.002	0.428	0.022	0.003
1,000	0.242	0.010	0.001	0.461	0.015	0.002
2,500	0.252	0.007	0.001	0.491	0.012	0.002
5,000	0.256	0.005	0.001	0.508	0.008	0.001
7,500	0.257	0.004	0.000	0.514	0.006	0.001
10,000	0.258	0.002	0.000	0.517	0.003	0.000

Figure 2-1

Histograms of *Heterobothrium okamotoi* count (HC) (**a**) and standard length (SL, cm) (**b**) with the estimated density.

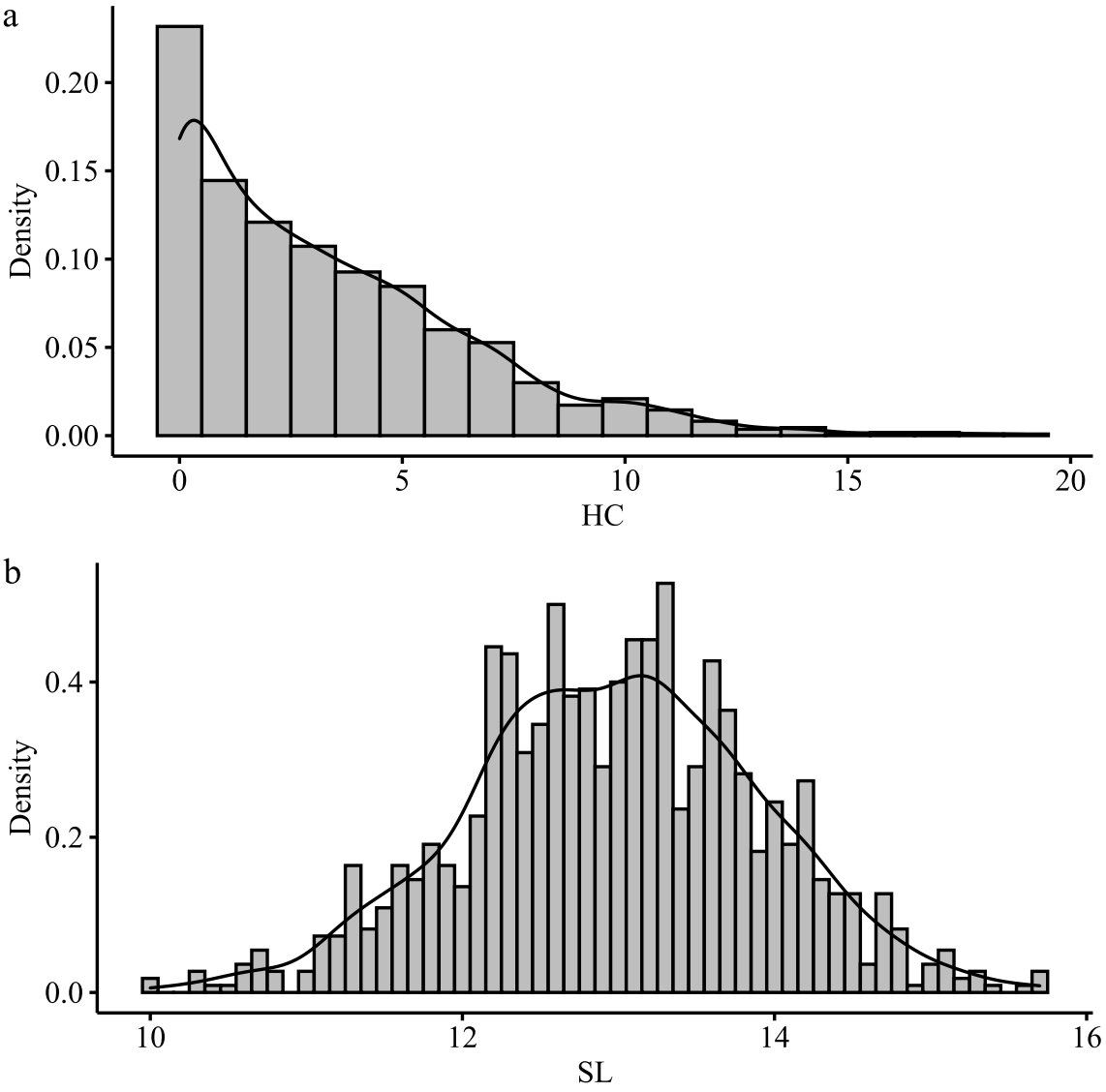

Figure 2-2

Population structure visualized by *t*-SNE plot of the genomic SNP data. Each filled circle represents one fish and its filled color shows *Heterobothrium okamotoi* count based on the color bar (right panel).

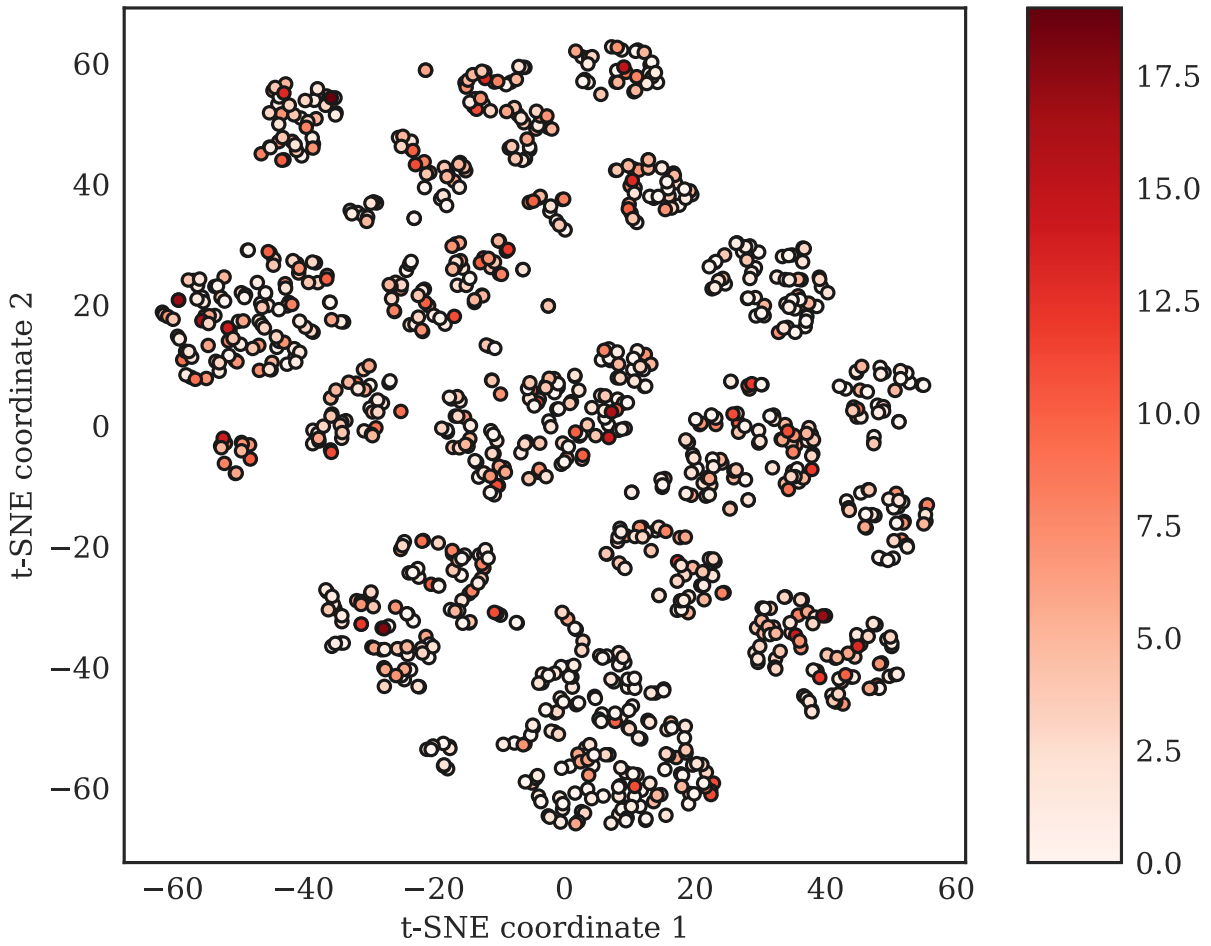

Figure 2-3

Manhattan plots of genome-wide association study with *Heterobothrium okamotoi* count (HC) (**a**) and standard length (SL) (**b**). The X-axis and Y-axis are the physical order of the SNP markers across the 22 chromosomes of *Takifugu rubripes* and the negative logarithm of *p*-values (base: 10) for the target trait, respectively. Each filled circle represents one SNP marker, and different colors of the circles distinguishes adjacent chromosomes. Red dashed lines are Bonferroni-corrected significance thresholds of 5.4 (= $-\log_{10}$ (0.05/12,584)).

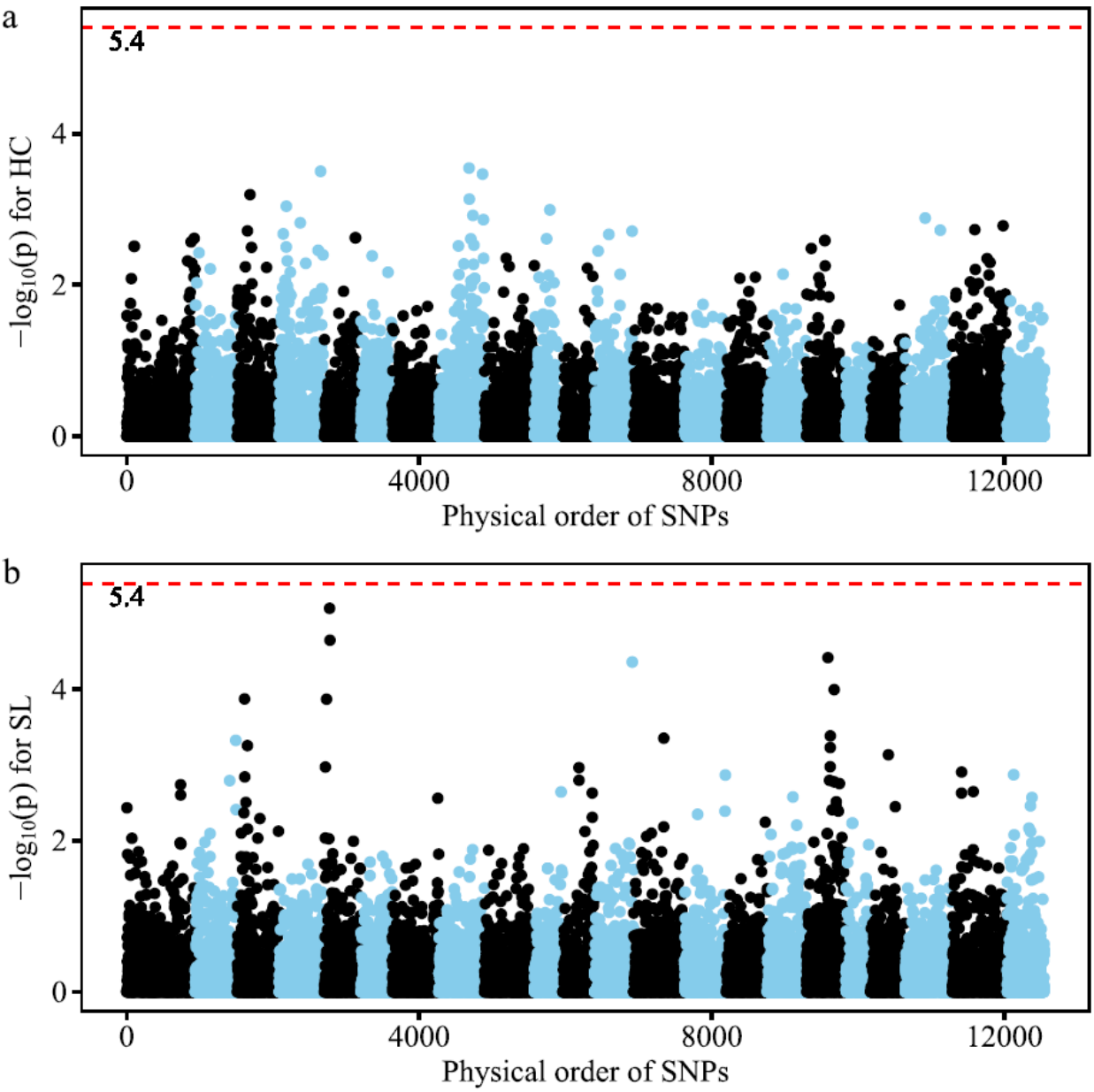

Figure 2-4

Scatter plots of estimated absolute marker effects for *Heterobothrium okamotoi* count (HC) (**a**) and standard length (SL) (**b**) under Bayes C model. The X-axis and Y-axis are the physical order of the SNP markers across the 22 chromosomes of *Takifugu rubripes* and the estimated absolute marker effects for the target trait, respectively. Each filled circle represents one SNP marker, and different colors of the circles distinguishes adjacent chromosomes.

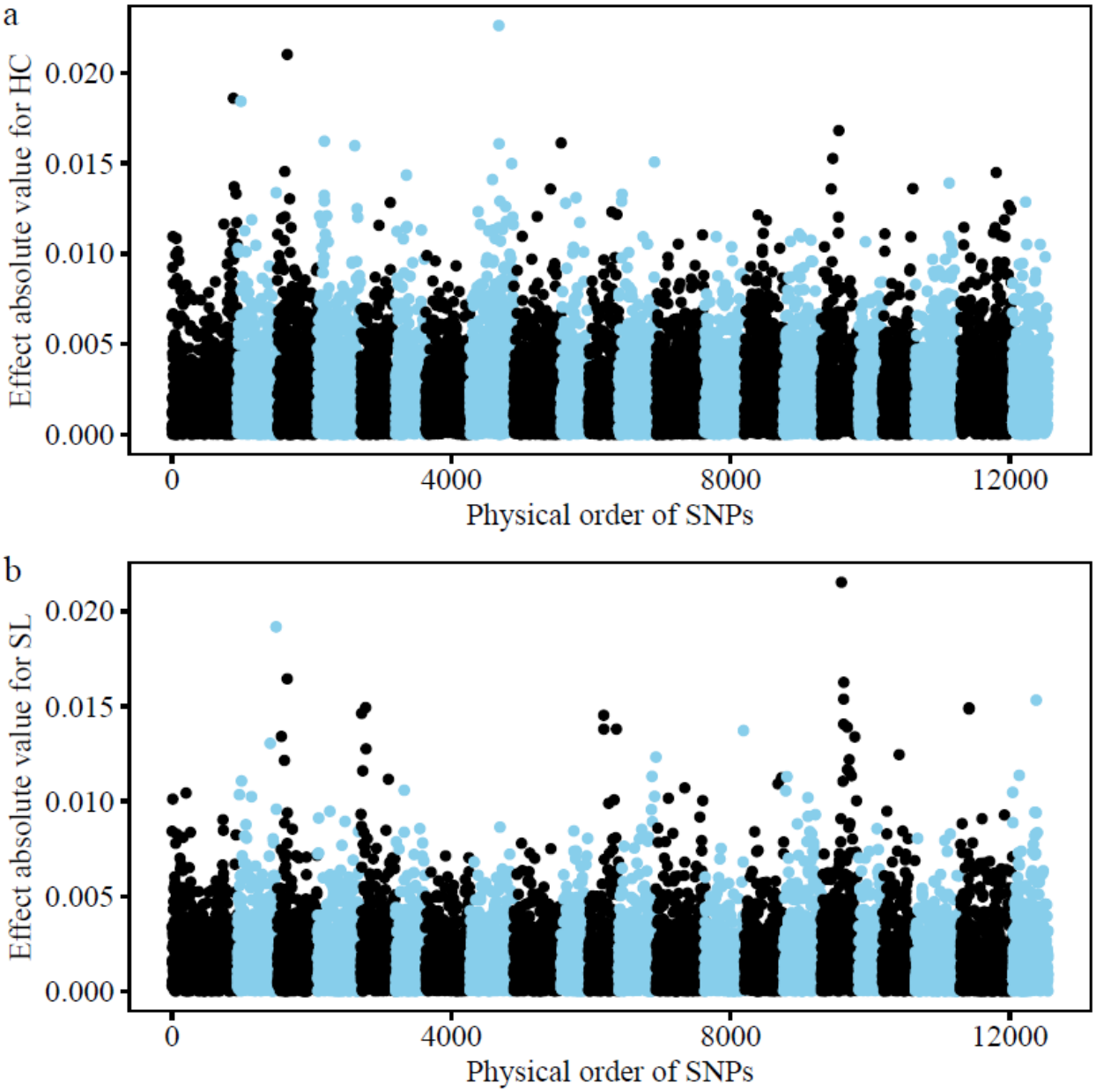

Figure 2-5

The box plot of predicitive abilies of genomic prediction for *Heterobothrium okamotoi* count (HC) (**a**) and standard length (SL) (**b**) using GBLUP model with various SNP density (50, 100, 500, 1,000, 2,500, 5,000, 7,500 and 10,000 SNPs). Red dashed line represents the average predictive ability of genomic prediction using the full SNP panel (12,548 SNPs).

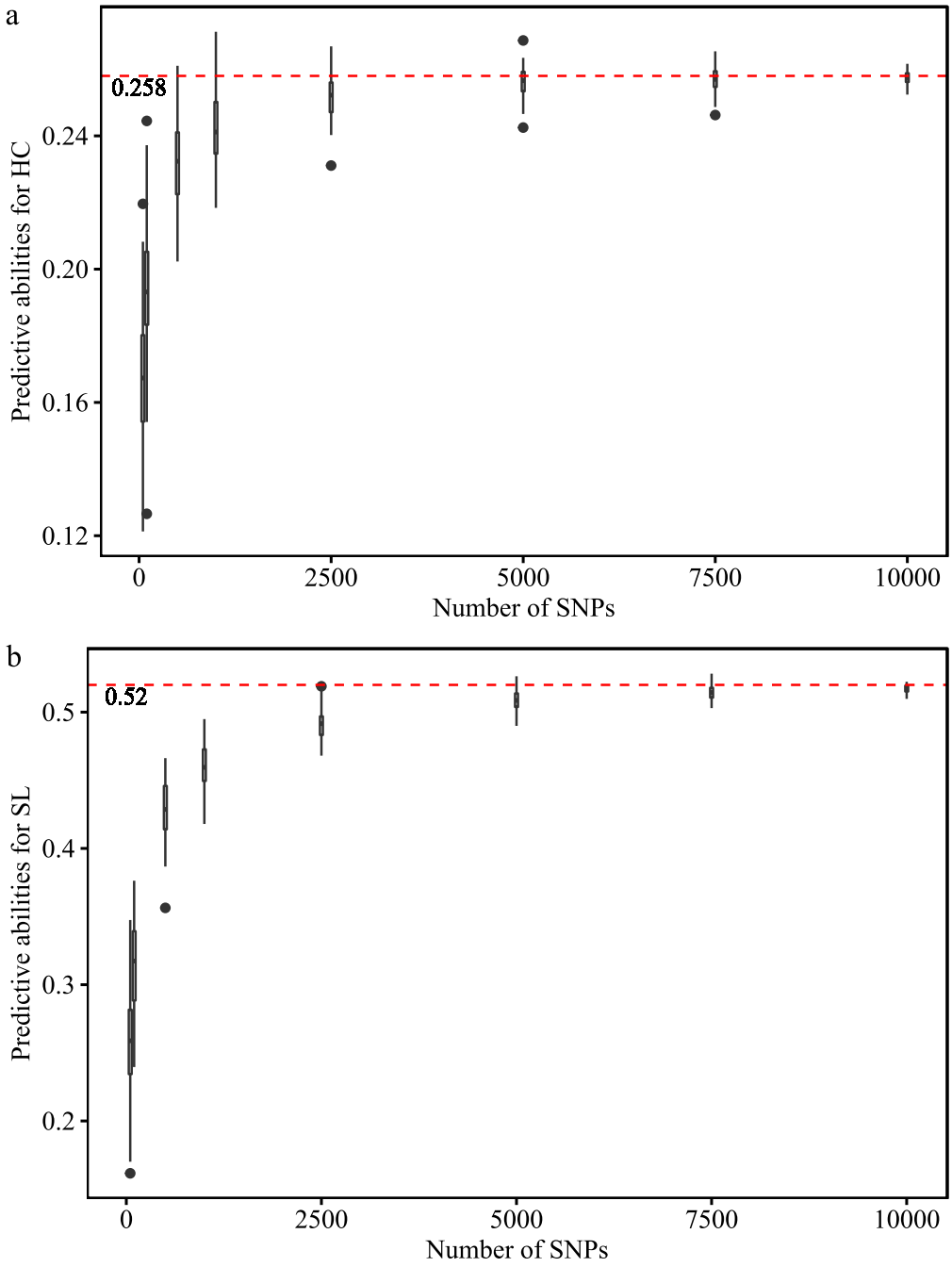

Table 3-1

The chemical composition, energy, amino acid composition, and polyunsaturated fatty acids (PUFAs) of the control and four experimental diets. FM: fishmeal, PP: plant protein, BAC: bacterial protein, LY: low level of yeast protein, HY: high level of yeast protein

	FM	PP	BAC	LY	HY
Fishmeal (%)	80	50	50	50	30
Replacement		plant	bacteria	yeast	yeast
Chemical composition (%)					
Moisture	6.43	5.73	6.72	3.94	5.5
Crude protein	55.37	55.53	53.66	57.11	54.62
Crude fat	6.93	8.02	8.23	7.42	8.55
Crude fiber	0.12	0.13	0.33	0.38	0.12
Crude ash	18.01	12.26	9.23	13.01	15.32
Nitrogen free extract	13.14	18.33	21.83	18.14	15.89
Energy (kcal/100g)					
Gross energy	437	471	478	475	461
Metabolizable energy	299	327	334	326	319
Amino acid composition (mg/100g)					
Isoleucine	2,210	2,370	2,470	2,240	2,180
Leucine	3,950	4,070	4,170	5,460	3,900
Lysine	4,070	4,210	4,330	3,260	3,710
Methionine	1,460	1,510	1,630	1,590	1,600
Cysteine	515	502	502	690	488
Phenylalanine	2,200	2,340	2,460	2,640	2,200
Tyrosine	1,710	1,860	1,960	2,050	1,780
Threonine	2,300	2,420	2,480	2,180	2,270
Tryptophan	617	686	709	554	770
Valine	2,620	2,780	2,900	2,570	2,690

Histidine	1,560	1,480	1,440	1,470	1,400
Arginine	3,250	3,170	3,130	2,890	3,230
Alanine	3,710	3,480	3,400	3,950	3,650
Aspartic acid	4,800	4,970	5,070	4,540	4,630
Glutamic acid	7,140	6,930	6,870	8,940	6,680
Glycine	4,140	3,500	3,240	3,360	3,660
Proline	2,470	2,310	2,240	3,360	2,350
Serine	2,170	2,350	2,470	2,470	2,060
Free Taurine	450	570	750	560	560
PUFAs (g/100g)	1.95	1.87	1.88	2.61	1.82

Table 3-2

Means ± S.D. of standard lengths (SL, cm), body weights (BW, g), and the growth rates of SL and BW during the three-month feeding trial (ΔSL_{41} & ΔBW_{41}, %) for the five treatment groups: FM, fishmeal; PP, plant protein; BAC, bacterial protein; LY, low level yeast protein; HY, high level yeast protein. SL_n (BW_n) represents the *n*-th measurement of SL (BW) during the feeding trial (n = 1, 2, 3, 4). Each group has 72 fish, except for FM which had one mort fish

	FM	PP	BAC	LY	HY
SL_1	19.31 ± 1.22	19.50 ± 1.22	18.95 ± 1.16	19.08 ± 1.27	19.42 ± 1.25
SL_2	19.60 ± 1.25	19.72 ± 1.22	19.36 ± 1.21	19.42 ± 1.27	19.80 ± 1.30
SL_3	20.48 ± 1.23	20.68 ± 1.22	20.14 ± 1026	20.32 ± 1.23	20.88 ± 1.32
SL_4	22.90 ± 1.37	23.07 ± 1.35	22.42 ± 1.35	22.73 ± 1.33	23.42 ± 1.41
BW_1	219.28 ± 39.12	226.56 ± 37.49	206.72 ± 35.77	212.47 ± 39.46	221.10 ± 41.41
BW_2	256.06 ± 45.35	258.26 ± 43.58	243.06 ± 38.71	247.74 ± 43.54	259.74 ± 48.39
BW_3	301.49 ± 52.98	304.42 ± 51.20	290.58 ± 48.48	297.35 ± 47.58	320.85 ± 59.45
BW_4	355.35 ± 64.33	354.11 ± 60.70	339.33 ± 56.05	357.68 ± 57.67	381.03 ± 63.86
ΔSL_{41}	18.71 ± 4.21	18.45 ± 4.23	18.37 ± 3.52	19.24 ± 4.13	20.68 ± 4.21
ΔBW_{41}	63.21 ± 19.40	57.23 ± 17.56	64.93 ± 13.57	70.44 ± 21.28	73.88 ± 19.06

Table 3-3

Akaike's information criterion (AIC), the difference between the tested and minimum AIC (ΔAIC), Akaike weight (*w*) of each candidate generalized linear model to compare SL_1, BW_1, ΔSL_{41}, ΔBW_{41} between each experimental and control groups

Test	Phenotype	Model	AIC	ΔAIC	*w*
Test 1	SL_1	Model 1	1172.8	2.2	0.224
		Model 2	1176.8	6.2	0.029
		Model 3	**1170.6**	**0.0**	**0.665**
		Model 4	1174.8	4.2	0.081
Test 2	BW_1	Model 1	3654.0	3.5	0.129
		Model 2	3657.4	6.9	0.023
		Model 3	**3650.6**	**0.0**	**0.734**
		Model 4	3654.3	3.7	0.114
Test 3	ΔSL_{41}	Model 1	2041.1	7.6	0.017
		Model 2	2042.8	9.3	0.007
		Model 3	**2033.5**	**0.0**	**0.734**
		Model 4	2036.3	2.7	0.186
		Model 5	2040.3	6.7	0.026
		Model 6	2044.7	11.1	0.003
		Model 7	2040.3	6.7	0.026
		Model 8	2044.7	11.1	0.003
Test 4	ΔBW_{41}	Model 1	3141.6	38.2	0.000
		Model 2	3129.5	26.1	0.000
		Model 3	3115.0	11.7	0.002
		Model 4	**3103.4**	**0.0**	**0.771**
		Model 5	3117.1	13.7	0.001
		Model 6	3107.2	3.9	0.112
		Model 7	3117.1	13.7	0.001
		Model 8	3107.2	3.9	0.112

The model with the highest w in each test is indicated in **BOLD**.

Model 1: phenotype = 1 + error, family = Gaussian

Model 2: phenotype = 1 + error, family = Gamma

Model 3: phenotype = diet + error, family = Gaussian

Model 4: phenotype = diet + error, family = Gamma

Model 5: phenotype = tank + error, family = Gaussian

Model 6: phenotype = tank + error, family = Gamma

Model 7: phenotype = diet + tank + error, family = Gaussian

Model 8: phenotype = diet + tank + error, family = Gamma

Table 3-4

Summary (mean ± S.D., n=30) of the blood chemistry analysis and liver weights in the five treatment groups: FM, fishmeal; PP, plant protein; BAC, bacterial protein; LY, low level yeast protein; HY, high level yeast protein

	FM	PP	BAC	LY	HY
Ht	29.3 ± 5.6	30.2 ± 3.4	29.5 ± 5.2	27.5 ± 4.5	25.2 ± 5.1
TCHO	206.8 ± 44.1	200.7 ± 33.4	175.9 ± 48.1	173.8 ± 36.0	179.6 ± 25.8
TG	109.1 ± 28.5	117.7 ± 27.2	113.3 ± 27.9	109.6 ± 23.3	118.9 ± 22.5
TP	5.00 ± 0.66	4.99 ± 0.45	5.17 ± 0.85	4.81 ± 0.54	4.69 ± 0.53
GLU	33.8 ± 11.5	31.8 ± 7.1	28.9 ± 5.0	29.4 ± 6.4	30.1 ± 4.8
LW	37.4 ± 9.0	34.6 ± 7.4	41.3 ± 8.0	39.6 ± 9.5	36.1 ± 8.1
HSI	10.3 ± 1.5	9.9 ± 1.4	12.1 ± 5.2	10.8 ± 2.0	9.8 ± 1.3

Ht: hematocrit (%), TCHO: total cholesterol (mg/dl), TG: triglyceride (mg/dl), TP: total protein (g/dl), GLU: glucose (mg/dl), LW: liver weight (g), HSI: hepatosomatic index (%)

Table 3-5

Akaike's information criterion (AIC), the difference between the tested and minimum AIC (ΔAIC), Akaike weight (*w*) of each candidate generalized linear model for hematocrit (Ht, %), total cholesterol (TCHO, mg/dl), triglyceride (TG, mg/dl), total protein (TP, g/dl), glucose (GLU, mg/dl), liver weight (LW, gram), and hepatosomatic index (HSI, %)

Test	Phenotype	Model	AIC	ΔAIC	*w*
Test 1	Ht	Model 1	916.5	12.2	0.001
		Model 2	920.2	15.9	0.000
		Model 3	**904.3**	**0.0**	**0.322**
		Model 4	908.3	3.9	0.045
		Model 5	904.6	0.3	0.275
		Model 6	908.5	4.1	0.041
		Model 7	904.6	0.3	0.275
		Model 8	908.5	4.1	0.041
Test 2	TCHO	Model 1	1536.6	10.5	0.005
		Model 2	1549.2	23.2	0.000
		Model 3	**1526.1**	**0.0**	**0.991**
		Model 4	1541.2	15.1	0.001
		Model 5	1539.1	13.0	0.001
		Model 6	1555.4	29.3	0.000
		Model 7	1539.1	13.0	0.001
		Model 8	1555.4	29.3	0.000
Test 3	TG	Model 1	1405.6	4.9	0.041
		Model 2	1410.2	9.5	0.004
		Model 3	1409.9	9.2	0.005
		Model 4	1414.8	14.1	0.000
		Model 5	1400.7	0.0	0.466
		Model 6	1408.6	7.9	0.009
		Model 7	**1400.7**	**0.0**	**0.466**
		Model 8	1408.6	7.9	0.009
Test 4	TP	Model 1	292.4	2.8	0.070
		Model 2	292.2	2.6	0.076
		Model 3	289.7	0.1	0.268
		Model 4	**289.6**	**0.0**	**0.280**
		Model 5	291.9	2.3	0.089
		Model 6	292.6	3.0	0.064
		Model 7	291.9	2.3	0.089
		Model 8	292.6	3.0	0.064
Test 5	GLU	Model 1	1033.1	31.8	0.000

		Model 2	1015.4	14.0	0.000
		Model 3	1032.4	31.1	0.000
		Model 4	1014.2	12.8	0.001
		Model 5	1019.8	18.4	0.000
		Model 6	1001.4	0.0	0.499
		Model 7	1019.8	18.4	0.000
		Model 8	**1001.4**	**0.0**	**0.499**
Test 6	LW	Model 1	1076.2	4.2	0.056
		Model 2	1075.9	3.9	0.064
		Model 3	**1072.0**	**0.0**	**0.456**
		Model 4	1072.2	0.2	0.409
		Model 5	1081.3	9.3	0.004
		Model 6	1082.0	10.0	0.003
		Model 7	1081.3	9.3	0.004
		Model 8	1082.0	10.0	0.003
Test 7	HSI	Model 1	629.9	40.3	0.000
		Model 2	613.5	23.8	0.000
		Model 3	607.8	18.1	0.000
		Model 4	**589.7**	**0.0**	**0.993**
		Model 5	620.2	30.5	0.000
		Model 6	601.1	11.4	0.003
		Model 7	620.2	30.5	0.000
		Model 8	601.1	11.4	0.003

The model with the highest w in each test is indicated in **BOLD**.

Model 1: phenotype = 1 + error, family = Gaussian

Model 2: phenotype = 1 + error, family = Gamma

Model 3: phenotype = diet + error, family = Gaussian

Model 4: phenotype = diet + error, family = Gamma

Model 5: phenotype = tank + error, family = Gaussian

Model 6: phenotype = tank + error, family = Gamma

Model 7: phenotype = diet + tank + error, family = Gaussian

Model 8: phenotype = diet + tank + error, family = Gamma

Table 3-6

Mean ± S.D. (n = 30) of the mature *Heterobothrium okamotoi* count (HC), *H. okamotoi* density (HD = 100 × HC/ SL_5^2), hematocrit (Ht, %), standard lengths and body weights at the end of feeding trial (SL_4 & BW_4) and parasite infection trial (SL_5 & BW_5); growth rates of standard lengths and body weights during the parasite infection trial (ΔSL_{54} and ΔBW_{54}, %) for the five treatment groups: FM, fishmeal; PP, plant protein; BAC, bacterial protein; LY, low level yeast protein; HY, high level yeast protein

	FM	PP	BAC	LY	HY
HC	23.5 ± 12.5	22.8 ± 18.9	24.0 ± 16.4	25.1 ± 11.7	25.2 ± 18.2
HD	3.9 ± 2.0	3.6 ± 2.9	4.1 ± 2.7	4.2 ± 1.9	3.8 ± 2.6
Ht	25.6 ± 3.0	27.2 ± 3.7	26.6 ± 5.2	24.6 ± 3.3	25.6 ± 3.9
SL_4	22.6 ± 1.5	23.4 ± 1.3	22.3 ± 1.5	22.2 ± 1.1	23.4 ± 1.4
SL_5	24.1 ± 1.3	25.1 ± 1.5	23.7 ± 1.4	24.4 ± 1.2	25.3 ± 1.6
BW_4	349.0 ± 60.9	361.3 ± 61.5	331.3 ± 58.8	340.9 ± 51.0	387.4 ± 67.4
BW_5	376.8 ± 58.1	396.8 ± 61.0	355.7 ± 61.0	375.7 ± 52.2	427.2 ± 73.7
ΔSL_{54}	6.9 ± 3.5	7.3 ± 2.5	6.6 ± 3.0	10.0 ± 2.6	7.8 ± 2.9
ΔBW_{54}	8.6 ± 7.3	10.2 ± 2.5	7.6 ± 5.7	10.5 ± 5.3	10.5 ± 6.8

Table 3-7

Akaike's information criterion (AIC), the difference between the tested and minimum AIC (ΔAIC), Akaike weight (*w*) of each generalized linear model (GLM) for *Heterobothrium okamotoi* count (HC), *H. okamotoi* density (HD = 100 × HC/ SL_{t5}^2), hematocrit (Ht, %), growth rate of standard length and body weight during the parasite infection trial (ΔSL_{54} and ΔBW_{54}, %)

Test	Phenotype	Model	AIC	ΔAIC	*w*
Test 1	HC	Model 1	2404.2	1153.3	0.000
		Model 2	1250.9	0.1	0.478
		Model 3	2406.6	1155.8	0.000
		Model 4	1258.6	7.7	0.010
		Model 5	2346.6	1095.8	0.000
		Model 6	**1250.9**	**0.0**	**0.499**
		Model 7	2349.0	1098.1	0.000
		Model 8	1258.2	7.4	0.013
Test 2	HD	Model 1	692.5	0.6	0.416
		Model 3	699.2	7.3	0.014
		Model 5	**691.9**	**0.0**	**0.551**
		Model 7	698.7	6.8	0.019
Test 3	Ht	**Model 1**	**840.8**	**0.0**	**0.224**
		Model 2	841.1	0.3	0.196
		Model 3	841.0	0.2	0.203
		Model 4	841.4	0.6	0.166
		Model 5	843.5	2.6	0.060
		Model 6	843.8	2.9	0.052
		Model 7	843.7	2.8	0.055
		Model 8	844.1	3.3	0.044
Test 4	ΔSL_{54}	Model 1	770.1	15.9	0.000
		Model 3	**754.2**	**0.0**	**0.978**
		Model 5	763.2	9.0	0.011
		Model 7	763.2	9.0	0.011
Test 5	ΔBW_{54}	Model 1	983.8	7.5	0.012
		Model 3	986.5	10.2	0.003
		Model 5	976.3	0.0	0.493
		Model 7	**976.3**	**0.0**	**0.493**

The model with the highest w in each test is indicated in **BOLD**.

Model 1: phenotype = 1 + error, family = Gaussian

Model 2: phenotype = 1 + error, family = Gamma

Model 3: phenotype = diet + error, family = Gaussian

Model 4: phenotype = diet + error, family = Gamma

Model 5: phenotype = tank + error, family = Gaussian

Model 6: phenotype = tank + error, family = Gamma

Model 7: phenotype = diet + tank + error, family = Gaussian

Model 8: phenotype = diet + tank + error, family = Gamma

For HC, GLMs assuming Gaussian and Gamma distributions were replaced by GLMs assuming Poisson and negative binomial distributions, respectively. "Tank" represents the infection tanks for tests 1, 2, and 3 and the feeding tanks for tests 4 and 5, respectively.

Table 3-8

Gene sets showing differential expression and the corresponding adjusted *p*-values from the over-representation analysis (ORA) and gene set enrichment analysis (GSEA) using Kyoto encyclopedia of genes and genomes (KEGG) pathways

Method	Comparison	KEGG pathway	Adjusted *p*-value
KEGG-ORA	PP-FM	Terpenoid backbone biosynthesis	0.004
	BAC-FM	Glycine, serine, and threonine metabolism	0.000
		Ribosome	0.000
		Histidine metabolism	0.000
		Glyoxylate and dicarboxylate metabolism	0.009
		Linoleic acid metabolism	0.015
		Tryptophan metabolism	0.015
		Biosynthesis of amino acids	0.015
		Biosynthesis of cofactors	0.015
		Ascorbate and aldarate metabolism	0.015
		Glycolysis / Gluconeogenesis	0.015
		Alanine, aspartate, and glutamate metabolism	0.015
		Ether lipid metabolism	0.015
		Protein processing in endoplasmic reticulum	0.018
		Pantothenate and CoA biosynthesis	0.023
		Cysteine and methionine metabolism	0.023
		Steroid hormone biosynthesis	0.023
		beta-Alanine metabolism	0.023
		Biosynthesis of unsaturated fatty acids	0.023
		Fatty acid elongation	0.033
		Drug metabolism - cytochrome P450	0.044
	LY-FM	Steroid biosynthesis	0.000
		Terpenoid backbone biosynthesis	0.001
	HY-FM	Steroid biosynthesis	0.000
		Terpenoid backbone biosynthesis	0.002
KEGG-GSEA	PP-FM	Ribosome	0.000
		Steroid biosynthesis	0.001
		DNA replication	0.002
		Tryptophan metabolism	0.006
		Mismatch repair	0.017
		Fatty acid degradation	0.020

	Cardiac muscle contraction	0.037
	PPAR signaling pathway	0.039
BAC-FM	Endocytosis	0.002
	Salmonella infection	0.002
	Homologous recombination	0.025
	Apoptosis	0.025
	Spliceosome	0.025
	Proteasome	0.025
	NOD-like receptor signaling pathway	0.026
	Fanconi anemia pathway	0.029
	Tight junction	0.029
	Cytokine-cytokine receptor interaction	0.029
	Phagosome	0.030
	Cell cycle	0.030
	Oxidative phosphorylation	0.044
	Mismatch repair	0.046
LY-FM	Ribosome	0.000
	Steroid biosynthesis	0.000
	Oxidative phosphorylation	0.003
	Mismatch repair	0.003
	DNA replication	0.013
	Tryptophan metabolism	0.026
	Fatty acid degradation	0.033
	Terpenoid backbone biosynthesis	0.033
HY-FM	Ribosome	0.000
	Steroid biosynthesis	0.000
	Tryptophan metabolism	0.000
	Glycine, serine, and threonine metabolism	0.000
	Butanoate metabolism	0.000
	Valine, leucine and isoleucine degradation	0.001
	Fatty acid degradation	0.001
	Histidine metabolism	0.002
	Alanine, aspartate, and glutamate metabolism	0.003
	Ascorbate and aldarate metabolism	0.004
	Biosynthesis of cofactors	0.004
	Peroxisome	0.008
	Folate biosynthesis	0.016
	Arginine and proline metabolism	0.016

Drug metabolism - other enzymes	0.016
Tyrosine metabolism	0.020
Glycerolipid metabolism	0.020
Starch and sucrose metabolism	0.020
beta-Alanine metabolism	0.020
Terpenoid backbone biosynthesis	0.020
Glyoxylate and dicarboxylate metabolism	0.022
Metabolism of xenobiotics by cytochrome P450	0.031
Glycolysis / Gluconeogenesis	0.031
Pentose and glucuronate interconversions	0.031
One carbon pool by folate	0.031
Propanoate metabolism	0.032
Pyruvate metabolism	0.032
Neuroactive ligand-receptor interaction	0.033
Purine metabolism	0.033
Drug metabolism - cytochrome P450	0.043

Figure 3-1

Box plots of the data obtained in the four measurements of standard length (SL) (**a**) and body weight (BW) (**b**) during the three-month feeding trial. The box plots display the median, two hinges, and two whiskers for the five dietary treatment groups; green, standard fishmeal (FM); purple, plant protein (PP); orange, bacterial protein (BAC); yellow, low level yeast protein (LY); blue, high level yeast protein (HY). The lower and upper hinges correspond to the first and third quartiles. Data beyond the end of the whiskers (maximum 1.5 interquartile range) are treated as "outlying" points.

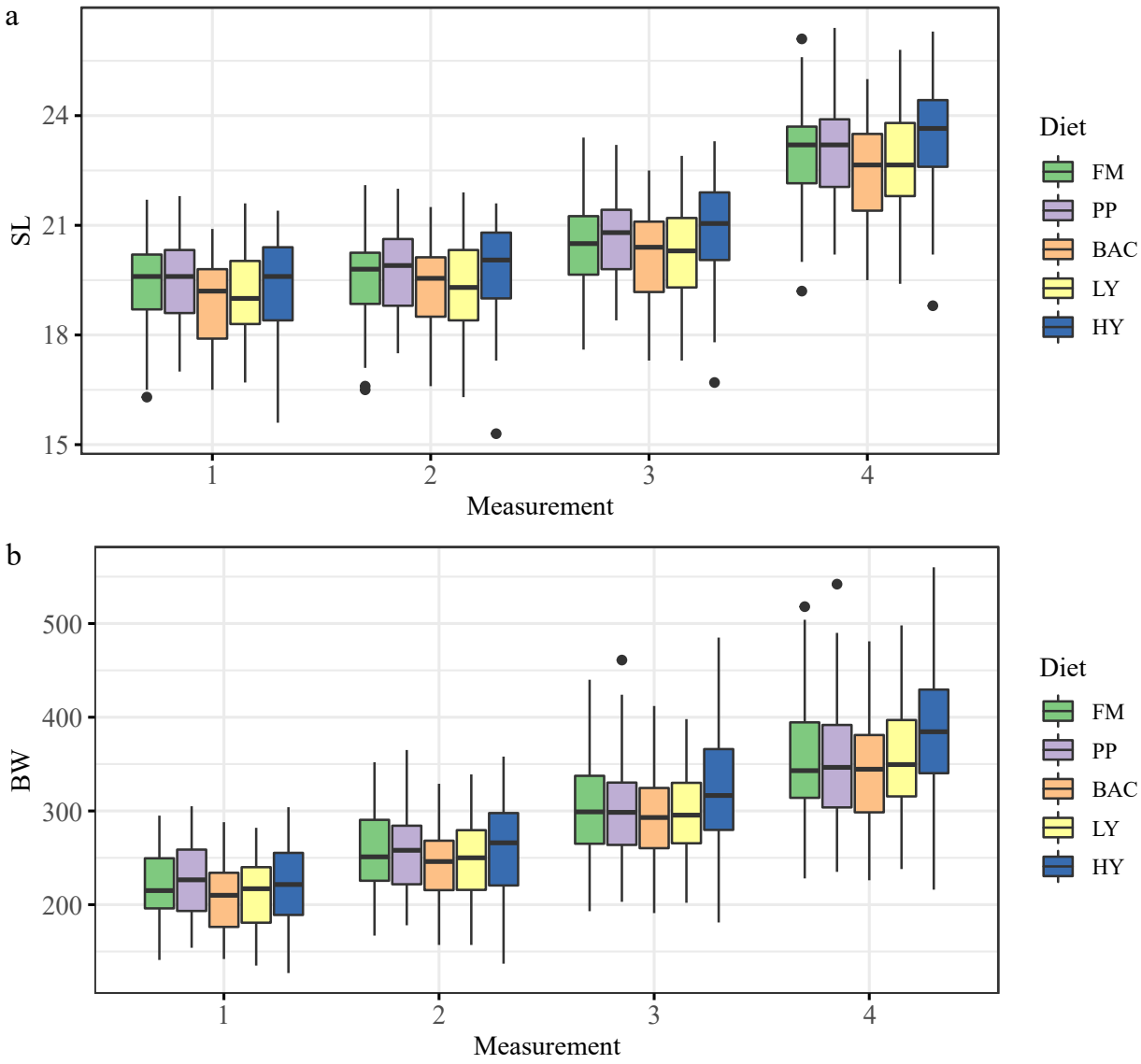

Figure 3-2

Plots of estimated marginal means (EMM) with 95% confidence intervals for (**a**) SL_1, (**b**) BW_1, (**c**) ΔSL_{41}, (**d**) ΔBW_{41} for the five treatment groups (FM, fishmeal; PP, plant protein; BAC, bacterial protein; LY, low level yeast protein; HY, high level yeast protein). Significant differences between the FM-fed group and the experimental group are marked by *p*-values from pair-wise *t*-tests ($\alpha = 0.05$).

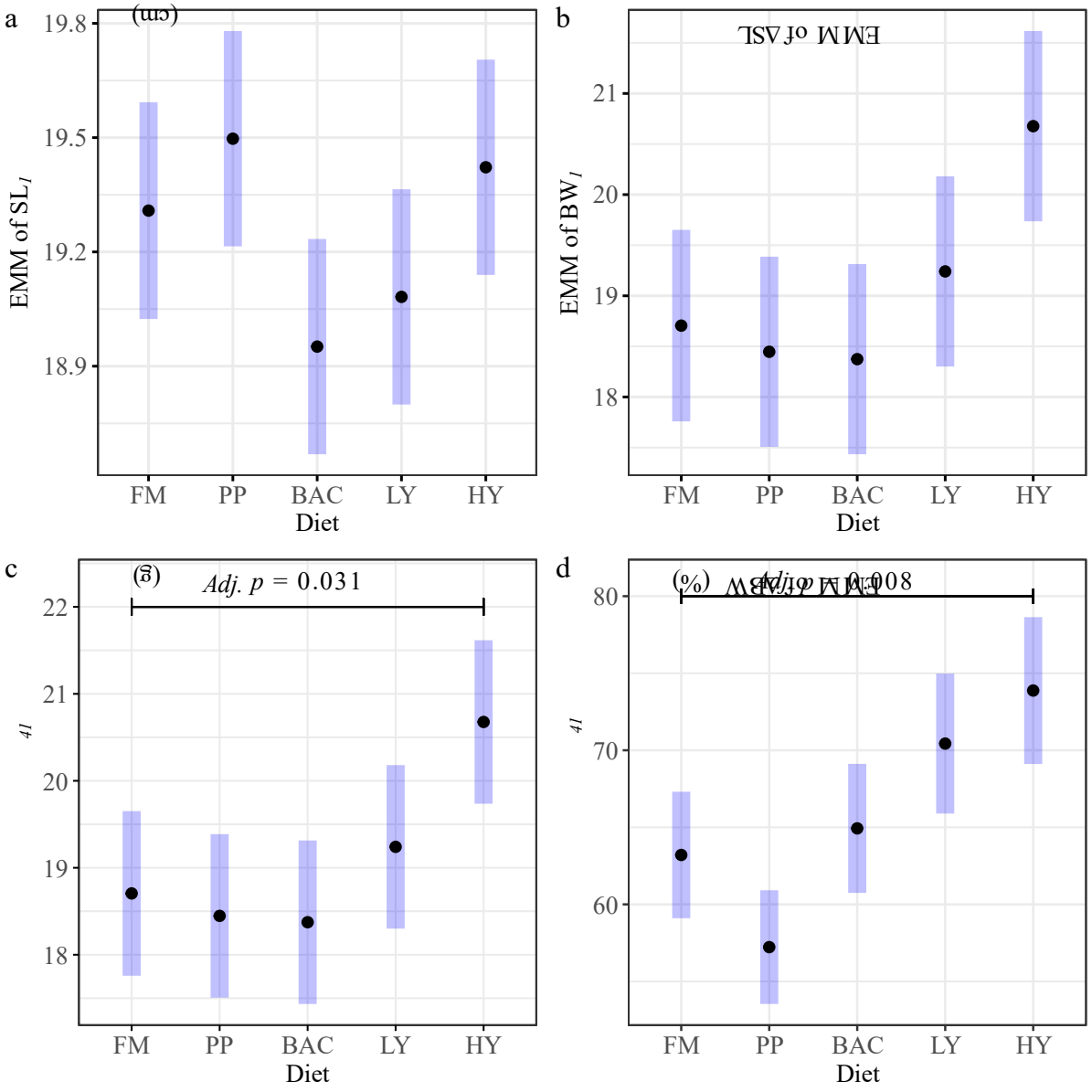

Figure 3-3

Box plots of the results from analyses of blood, liver weight, and the hepatosomatic index: (**a**) hematocrit (Ht, %), (**b**) total cholesterol (TCHO, mg/dl), (**c**) triglyceride (TG, mg/dl), (**d**) total protein (TP, g/dl), (**e**) glucose (GLU, mg/dl), (**f**) liver weight (LW, g), and (**g**) the hepatosomatic index (HSI, %). Each treatment group contained 30 fish (FM, fishmeal; PP, plant protein; BAC, bacterial protein; LY, low level yeast protein; HY, high level yeast protein). The box plots show the median, two hinges, and two whiskers. The lower and upper hinges correspond to the first and third quartiles. Data beyond the end of the whiskers (maximum 1.5 interquartile range) are treated as "outlying" points.

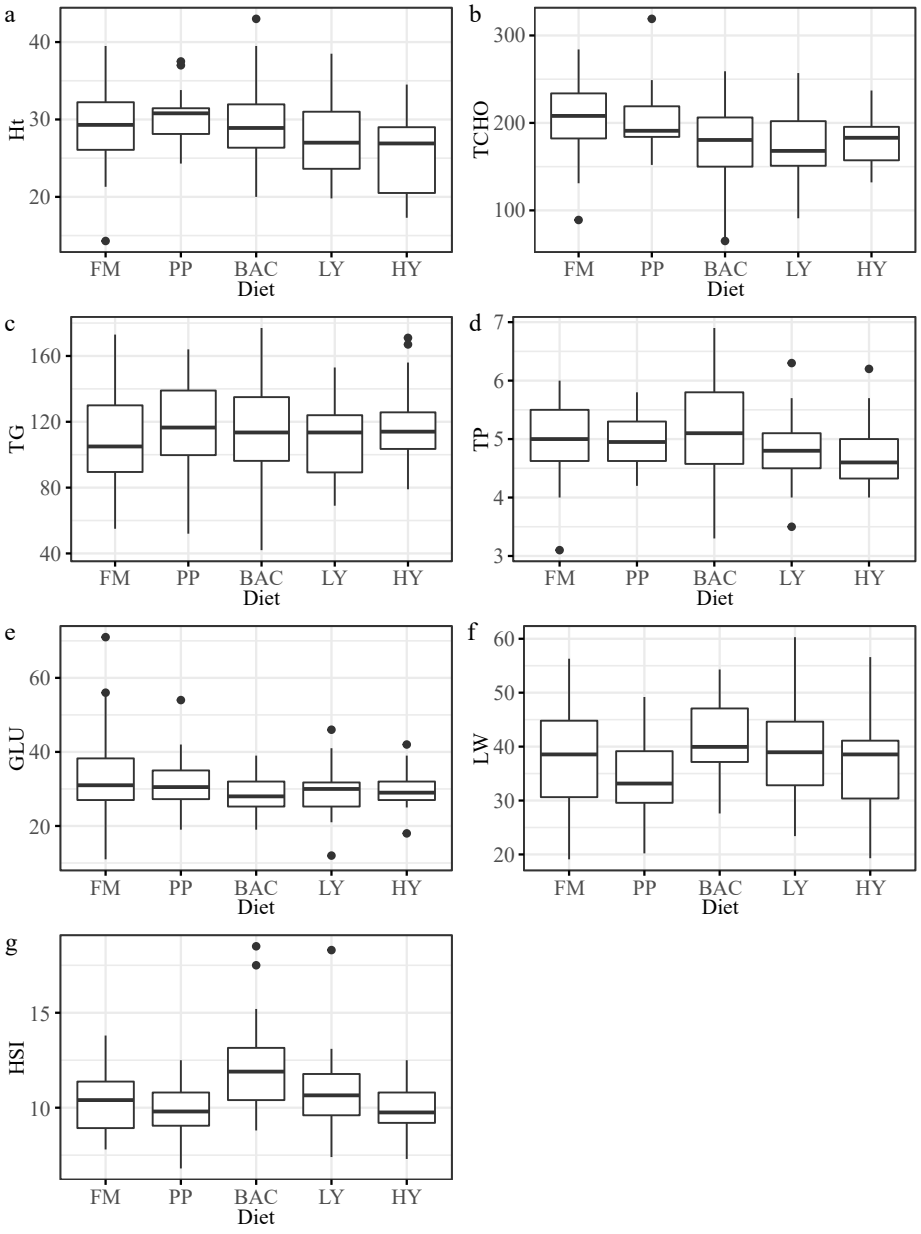

Figure 3-4

Plots of estimated marginal means (EMM) with 95% confidence intervals for (**a**) hematocrit (Ht, %), (**b**) total cholesterol (TCHO, mg/dl), (**c**) triglyceride (TG, mg/dl), (**d**) total protein (TP, g/dl), (**e**) glucose (GLU, mg/dl), (**f**) liver weight (LW, gram), and (**g**) hepatosomatic index (HSI, %) for the five treatment groups (FM, fishmeal; PP, plant protein; BAC, bacterial protein; LY, low level yeast protein; HY, high level yeast protein). Significant differences between the control and the experimental group are marked with *p*-values from pair-wise *t*-tests (α = 0.95).

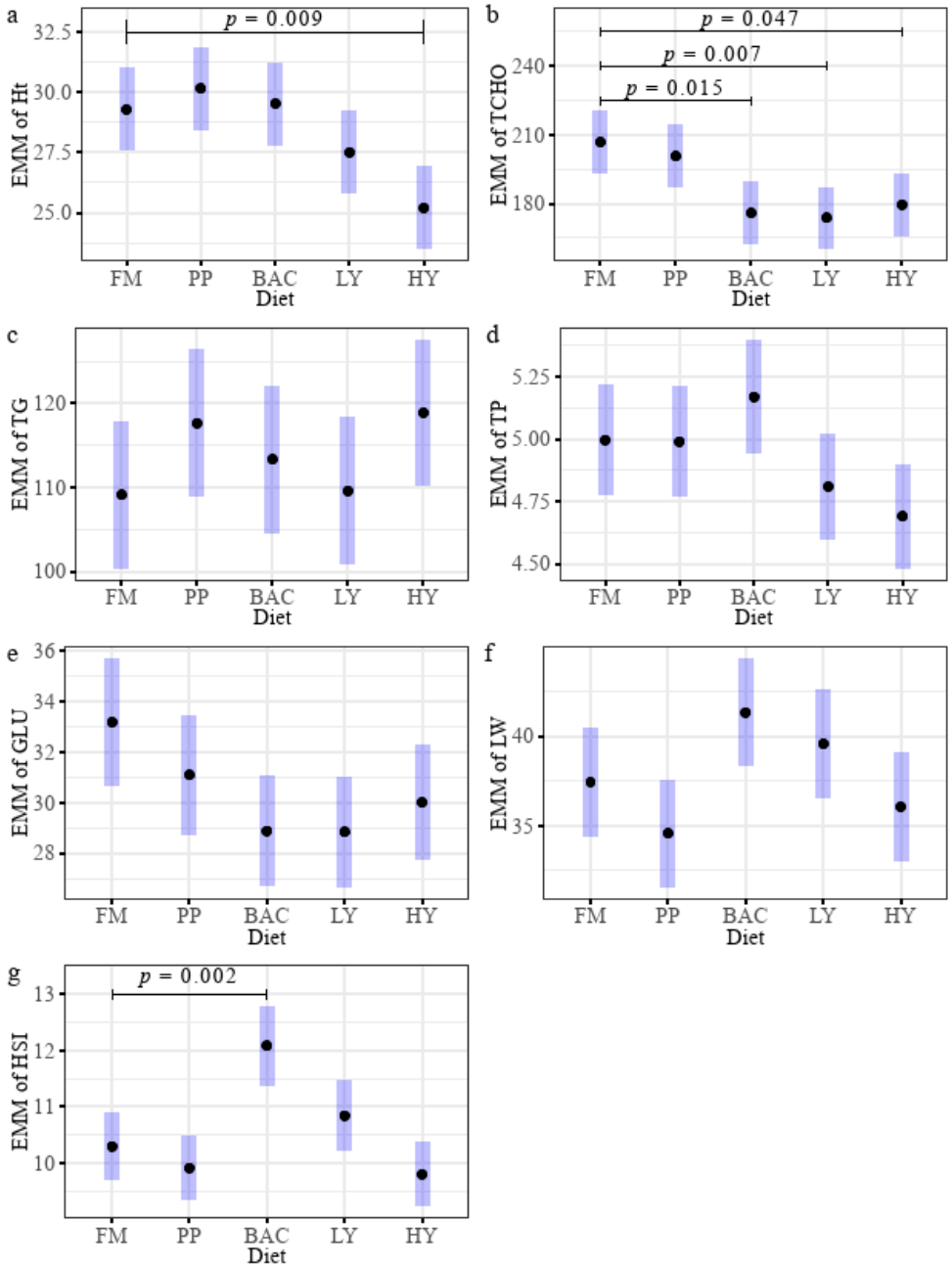

Figure 3-5

Box plots of (**a**) *Heterobothrium okamotoi* counts (HC), (**b**) *H. okamotoi* density (HD), (**c**) hematocrit (Ht, %), (**d**) (**e**) changes in standard lengths and body weights during the parasite infection trial (ΔSL_{54} and ΔBW_{54}, %) in the five diet groups (FM, fishmeal; PP, plant protein; BAC, bacterial protein; LY, low level yeast protein; HY, high level yeast protein). The box plots show the median, two hinges, and two whiskers. The lower and upper hinges correspond to the first and third quartiles. Data beyond the end of the whiskers (maximum 1.5 interquartile range) are treated as "outlying" points.

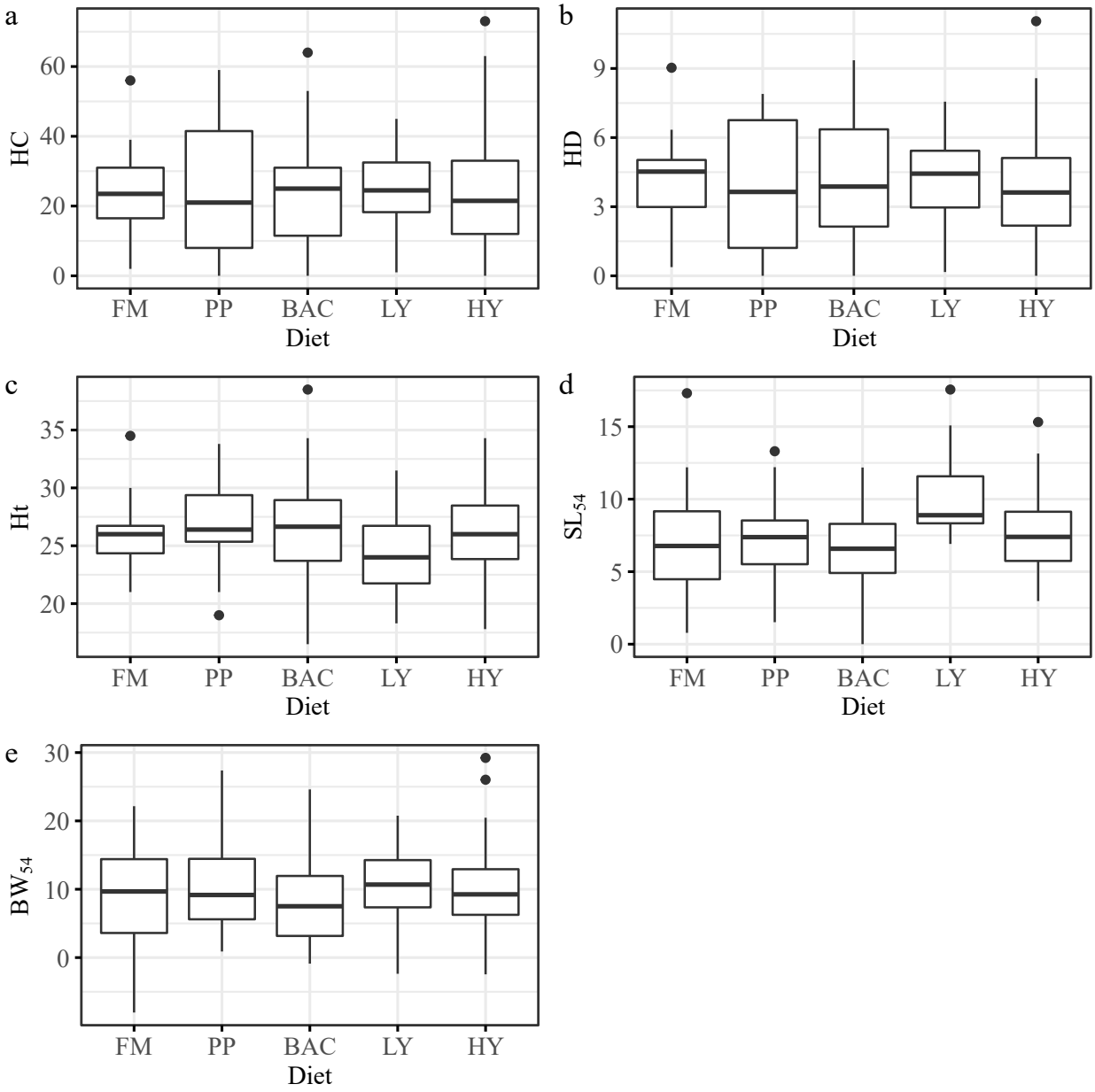

Figure 3-6

(**a**) Principal component analysis (PCA) and (**b**) *t*-distributed stochastic neighbor embedding (*t*-SNE) plot of pre-proceed $\log_2$(count per million) values with samples colored by sequencing libraries.

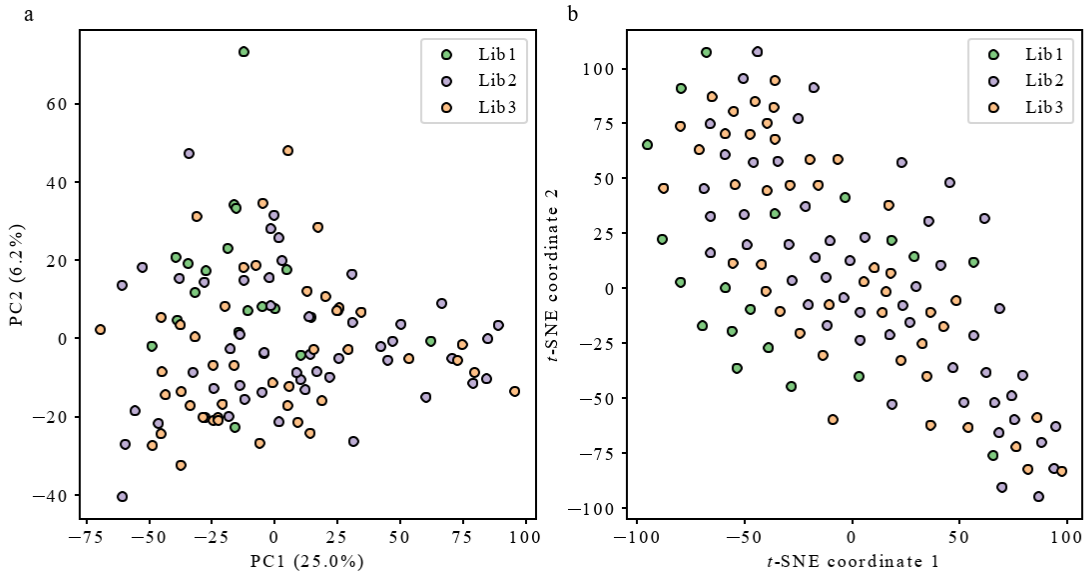

Figure 3-7

Venn diagram of the numbers of differentially expressed genes from the empirical Bayes moderated t-statistics among the four comparisons, namely, PP-FM (comparison between plant protein and fishmeal diets, purple), BAC-FM (bacterial protein and fishmeal diets, orange), LY-FM (low level yeast protein and fishmeal diets, yellow), and HY-FM (high level yeast protein and fishmeal diets, blue).

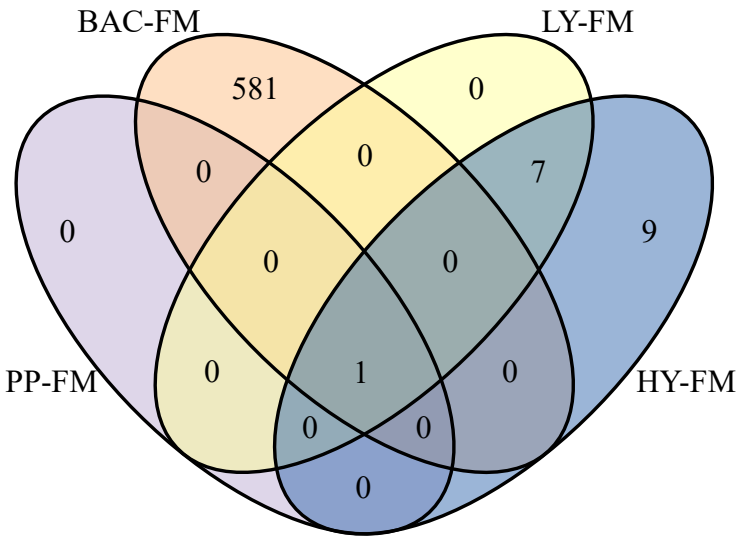

Figure 3-8

Bar charts illustrating gene counts and adjusted *p*-values (Adj. *p*) from the KEGG-based gene set enrichment analysis among four comparisons: (**a**) PP-FM (comparison between plant protein and fishmeal diets); (**b**) BAC-FM (bacterial protein and fishmeal diets); (**c**) LY-FM (low level yeast protein and fishmeal diets); (**d**) HY-FM (high level yeast protein and fishmeal diets).

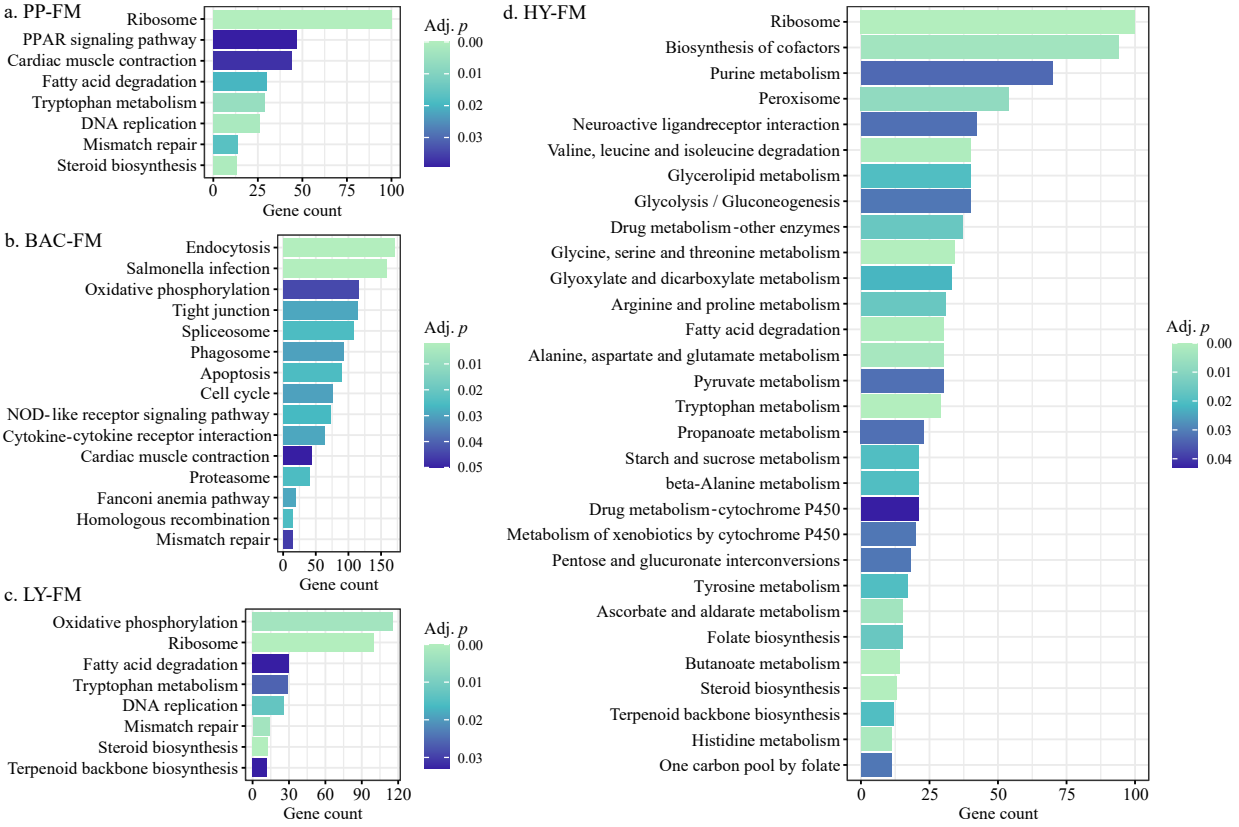

Figure 3-9

Bar charts showing gene counts and adjusted *p*-values (Adj. *p*) from the over-representation analysis of the Kyoto encyclopedia of genes and genomes (KEGG) pathways among the four comparisons: (**a**) PP-FM (comparison between plant protein and fishmeal diets); (**b**) BAC-FM (bacterial protein and fishmeal diets); (**c**) LY-FM (low level yeast protein and fishmeal diets); (**d**) HY-FM (high level yeast protein and fishmeal diets).

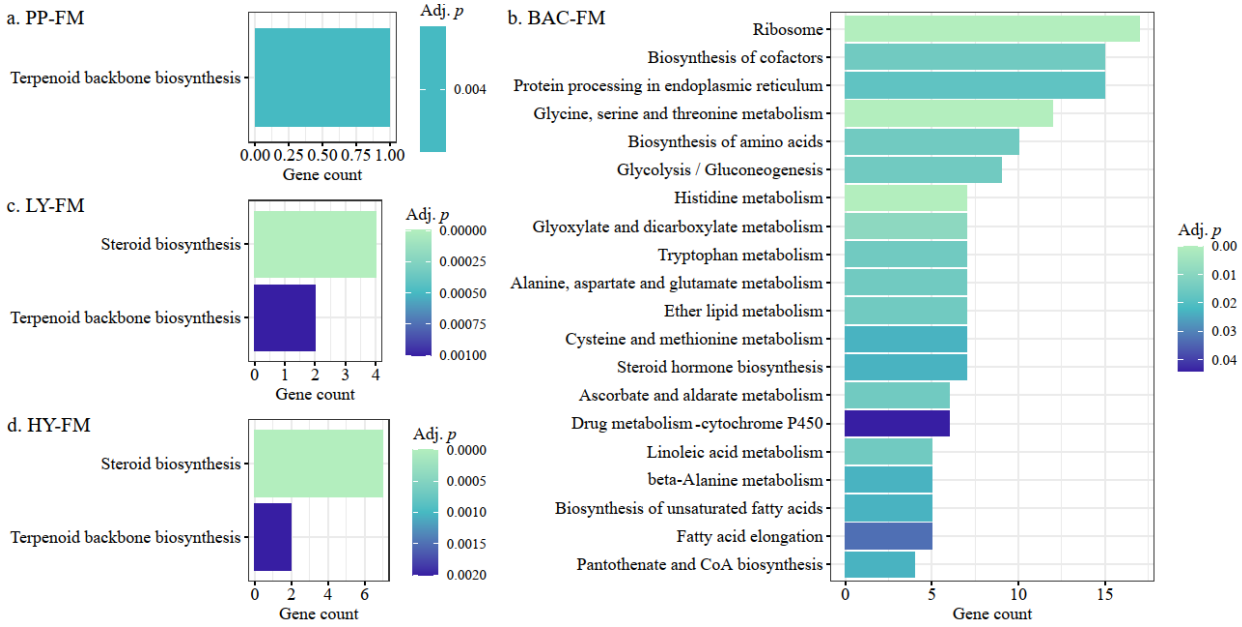

Table 4-1.

Predictive ability (mean ± standard error) of genomic prediction for *Heterobothrium okamotoi* count (HC) and standard length (SL) under three models, i.e., GBLUP, Bayes C, and Bayes reproducing kernel Hilbert space (Bayes RKHS) in Pop-C and Pop-CW which experienced a short-term dietary treatment of a low fish meal diet consisting of high proportion of yeast protein (HY)

Model	Pop-C		Pop-CW	
	HC	SL	HC	SL
GBLUP	0.520 ± 0.007	0.564 ± 0.006	0.602 ± 0.007	0.736 ± 0.003
Bayes C	0.521 ± 0.007	0.565 ± 0.006	0.601 ± 0.007	0.736 ± 0.003
Bayes RKHS	0.522 ± 0.007	0.567 ± 0.006	0.604 ± 0.007	0.740 ± 0.003

Figure 4-1

Histograms of *Heterobothrium okamotoi* count (HC) (**a**), transformed HC (**b**), and standard length (SL, cm) (**b**) with the estimated density for Pop-C, and those for Pop-CW (**d**, **e**, **f**).

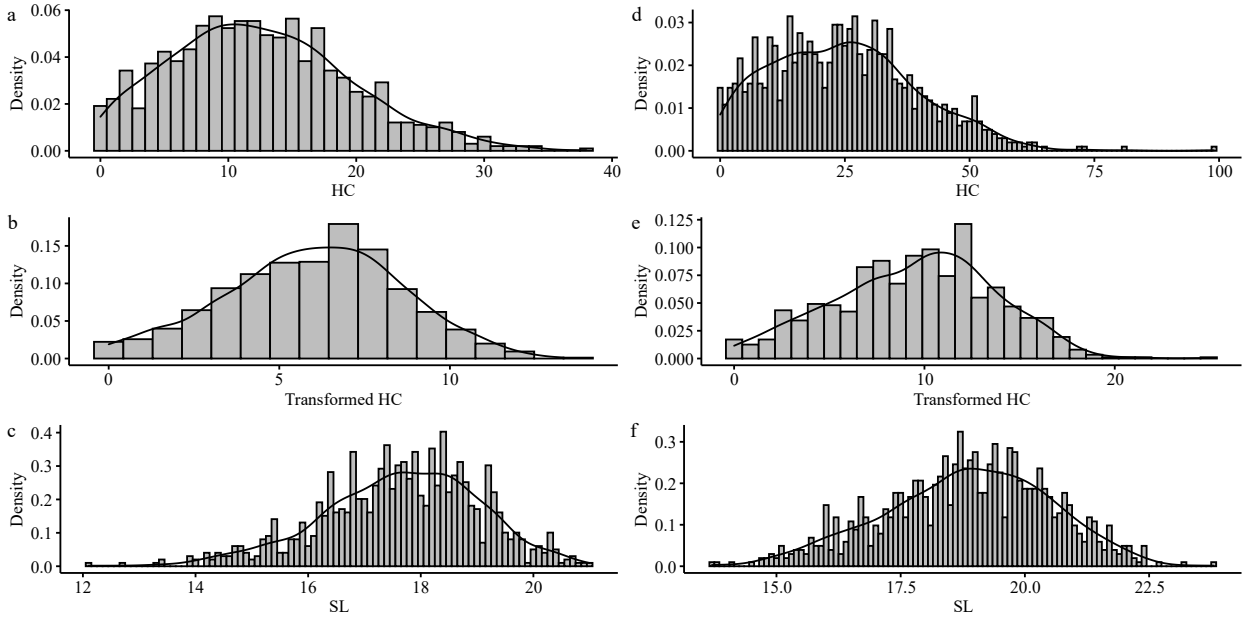

Figure 4-2

Population structure visualized by *t*-SNE plot of the genomic SNP data for Pop-C (**a**) and Pop-CW (**b**). Each filled circle represents one fish and its filled color shows *Heterobothrium okamotoi* count based on the color bar (right panel).

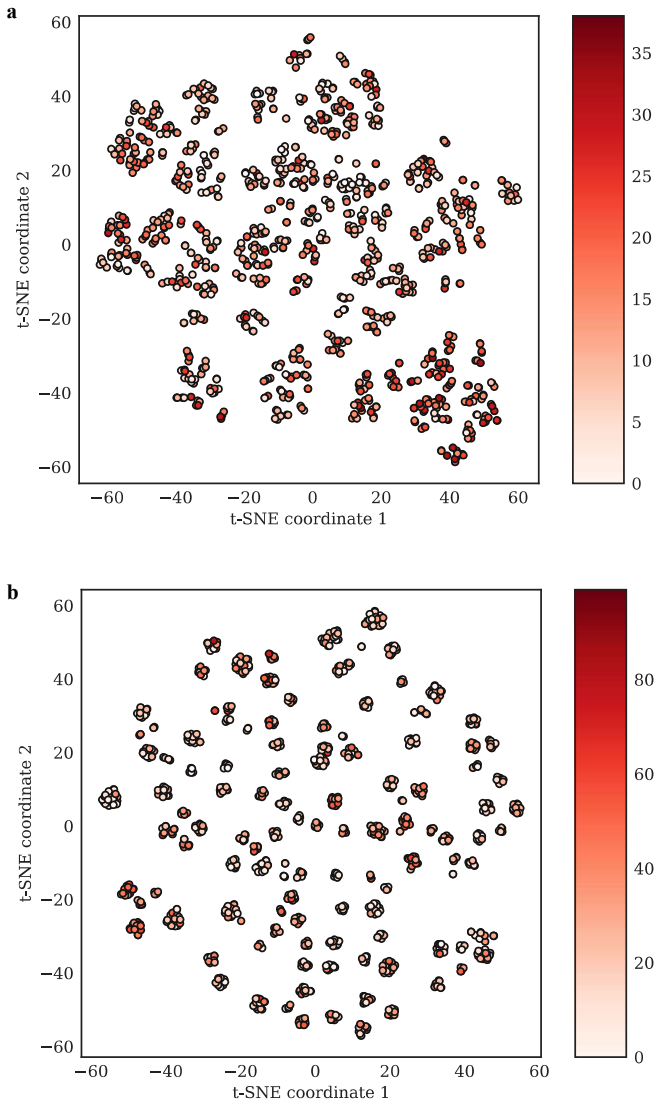

Figure 4-3

Manhattan plots of genome-wide association study with *Heterobothrium okamotoi* count (HC) (**a**) and standard length (SL) (**b**) for Pop-C and those for Pop-CW (**c**, **d**). Both populations experienced a short-term dietary treatment of a low fish meal diet including high proportion of yeast protein. The X-axis and Y-axis are the physical order of the SNP markers across the 22 chromosomes of *Takifugu rubripes* and the negative logarithm of *p*-values (base: 10) for the target trait, respectively. Each filled circle represents one SNP marker, and different colors of the circles distinguishes adjacent chromosomes. Red dashed lines are Bonferroni-corrected significance thresholds of 5.585 (= – $\log_{10}$ (0.05/19,227)) and 5.629 (= – $\log_{10}$ (0.05/21,272)) for Pop-C and Pop-CW, respectively. One significant SNP (Chr3: 5,531,819, *ube4b*) was observed for HC in Pop-C as shown in (**a**).

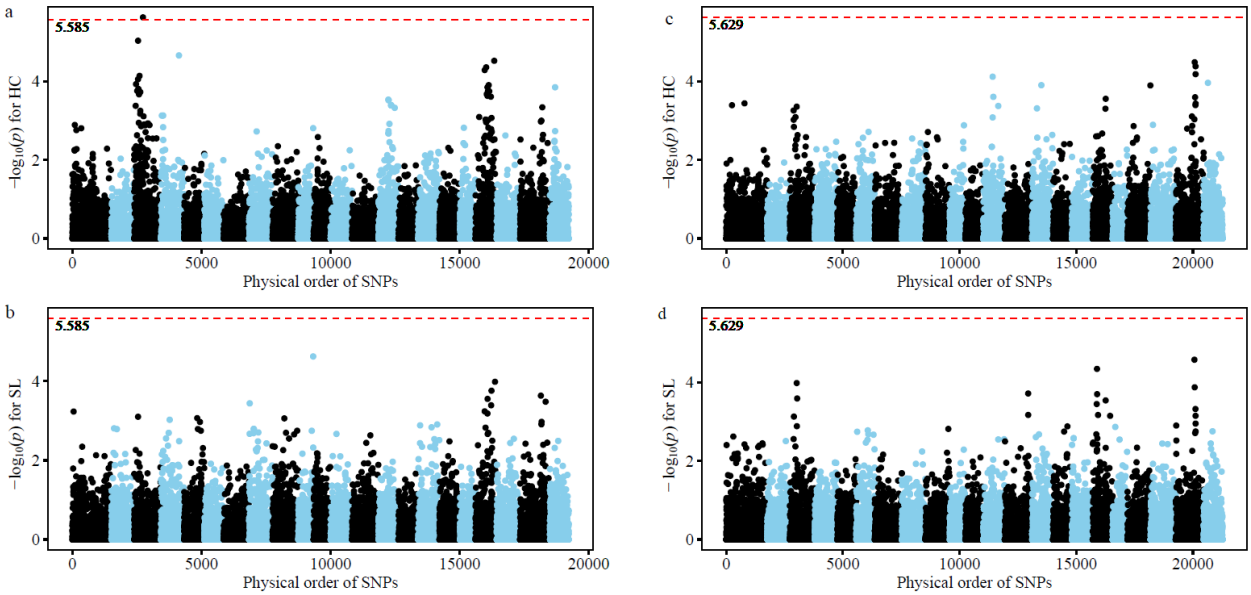

Figure 4-4

Scatter plots of estimated absolute marker effects for *Heterobothrium okamotoi* count (HC) (**a**) and standard length (SL) (**b**) and those for Pop-CW (**c**, **d**) under Bayes C model. The X-axis and Y-axis are the physical order of the SNP markers across the 22 chromosomes of *Takifugu rubripes* and the estimated absolute marker effects for the target trait, respectively. Each filled circle represents one SNP marker, and different colors of the circles distinguishes adjacent chromosomes.

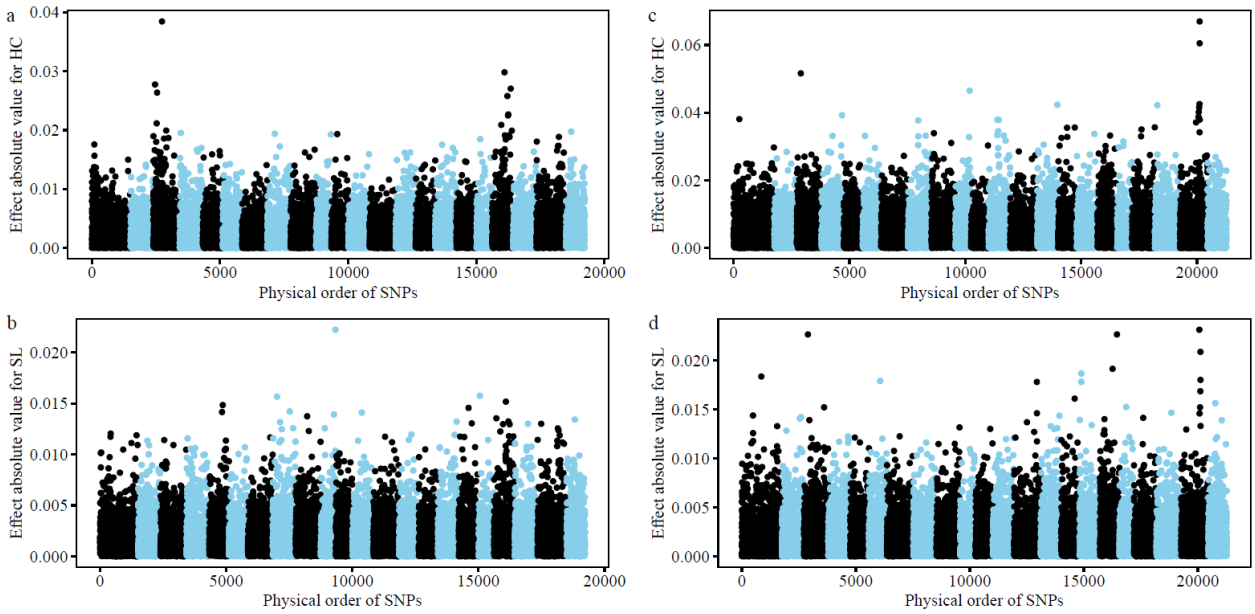

Table 5-1

Pearson's *r* between phenotypes for standard length (SL, cm) measured at different time point during a long-term feeding treatment: SL_1 (198th day), SL_2 (319th day), SL_3 (422nd day), and SL_4 (569th day). Upper triangular part of the table is estimated Pearson's *r* and the lower triangular part is *p* value

	SL_1	**SL_2**	**SL_3**	**SL_4**
SL_1	-	0.947	0.890	0.849
SL_2	$p < 0.001$	-	0.937	0.890
SL_3	$p < 0.001$	$p < 0.001$	-	0.946
SL_4	$p < 0.001$	$p < 0.001$	$p < 0.001$	-

Table 5-2

Genetic correlation between phenotypes for standard length (SL, cm) measured at different time point during a long-term feeding treatment: SL_1 (198th day), SL_2 (319th day), SL_3 (422nd day), and SL_4 (569th day). Upper triangular part of the table is estimated genetic correlation under a multivariate model

	SL_1	**SL_2**	**SL_3**	**SL_4**
SL_1	-	0.974	0.942	0.919
SL_2	-	-	0.959	0.917
SL_3	-	-	-	0.962
SL_4	-	-	-	-

Table 5-3

Predictive ability (mean ± standard error) on standard length (SL, cm) measured at different time point during a long-term feeding treatment: SL_1 (198th day), SL_2 (319th day), SL_3 (422nd day), and SL_4 (569th day) under three models: GBLUP, Bayes C, Bayes reproducing kernel Hilbert space (Bayes RKHS), in Pop-C and Pop-CW which experienced a short-term dietary treatment of a low fish meal diet including high proportion of yeast protein (HY)

Model	**SL_1**	**SL_2**	**SL_3**	**SL_4**
GBLUP	0.587 ± 0.006	0.643± 0.005	0.625 ± 0.006	0.627 ± 0.006
Bayes C	0.588 ± 0.006	0.643± 0.005	0.625 ± 0.006	0.627 ± 0.006
Bayes RKHS	0.589 ± 0.006	0.643± 0.005	0.630 ± 0.006	0.632 ± 0.006

Table 5-4

Pearson's *r* between genomic estimated breeding values (GEBVs) for standard length (SL, cm) measured at different time point during a long-term feeding treatment: SL_1 (198th day), SL_2 (319th day), SL_3 (422nd day), and SL_4 (569th day). Upper triangular part of the table is estimated Pearson's *r* and the lower triangular part is *p* value. GEBVs were calculated using GBLUP model

	SL_1	**SL_2**	**SL_3**	**SL_4**
SL_1	-	0.966	0.933	0.915
SL_2	$p < 0.001$	-	0.962	0.932
SL_3	$p < 0.001$	$p < 0.001$	-	0.962
SL_4	$p < 0.001$	$p < 0.001$	$p < 0.001$	-

Figure 5-1

Histograms with the estimated density of standard length (cm) at four different time point: SL_1 (198th day, a), SL_2 (319th day, b), SL_3 (422nd day, c), and SL_4 (569th, d).

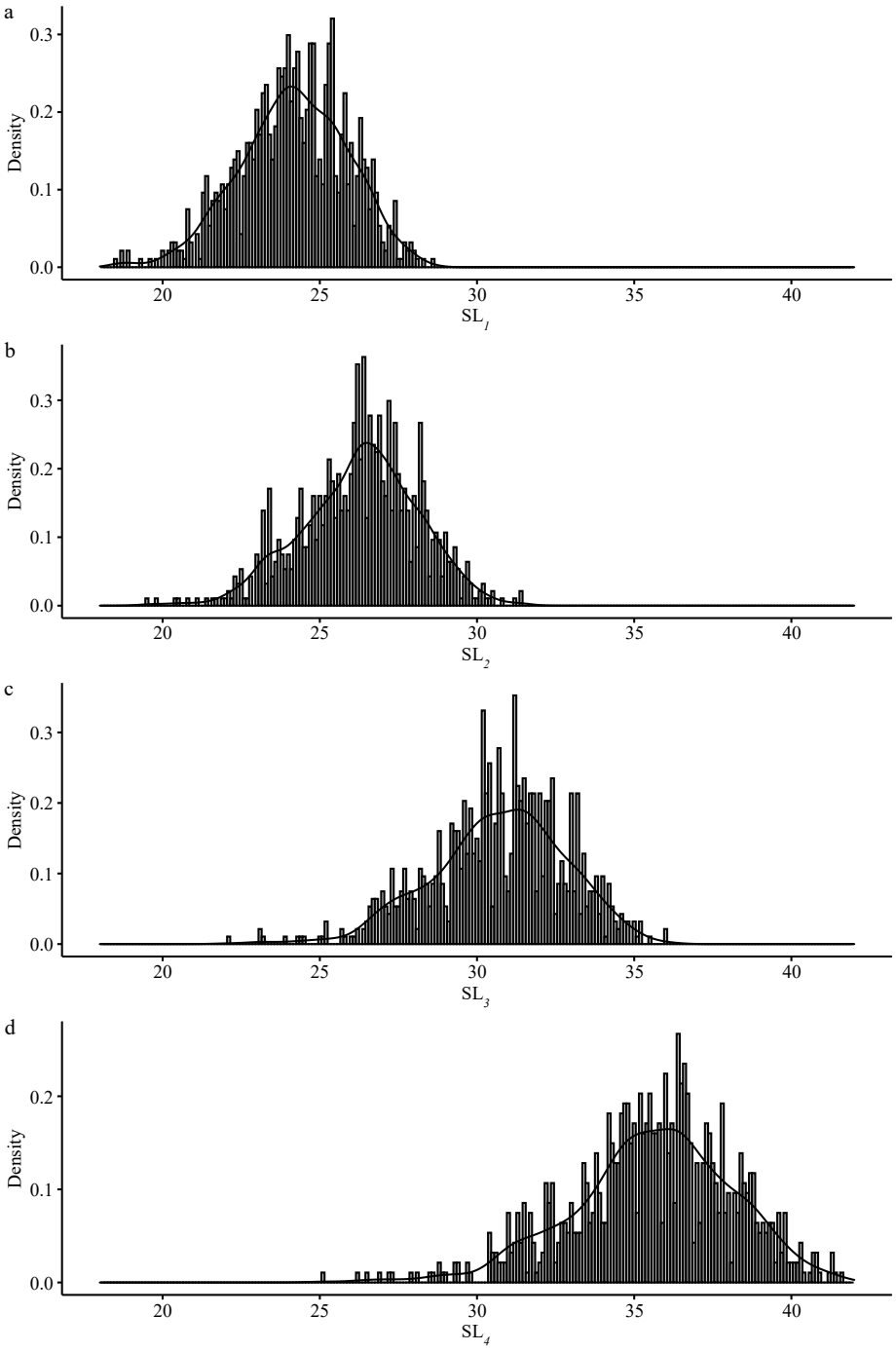

Figure 5-2

Population structure visualized by *t*-SNE plot of the genomic SNP data. Each circle represents one fish, and its filled color shows standard length (cm) measured at the 569th day of a long-term feeding treatment (SL_4) based on the color bar (right panel).

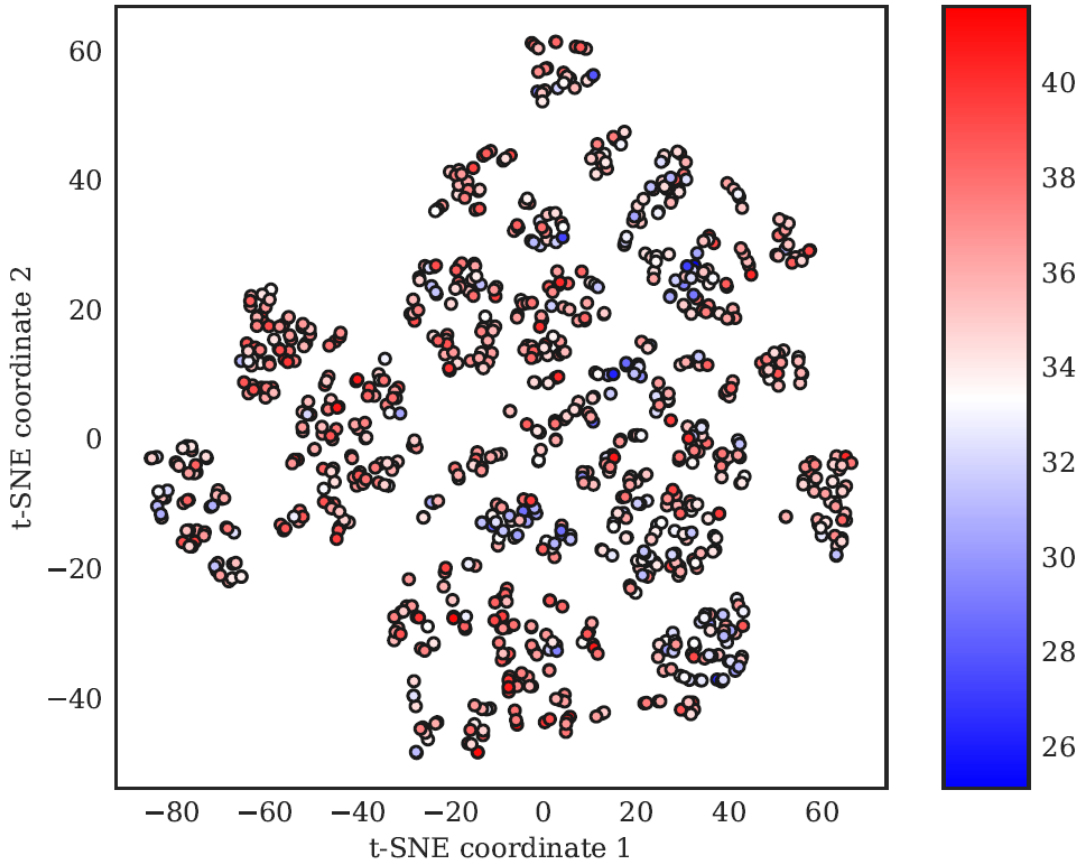

Figure 5-3

Manhattan plots of genome-wide association study with standard length at four different time points: SL_1 (a), SL_2 (b), SL_3 (c), and SL_4 (d). The X-axis and Y-axis are the physical order of the SNP markers across the 22 chromosomes of *Takifugu rubripes* and the negative logarithm of *p*-values (base: 10) for the target trait, respectively. Each filled circle represents one SNP marker, and different colors of the circles distinguishes adjacent chromosomes. Red dashed lines are Bonferroni-corrected significance thresholds of 5.518 ($= -\log_{10}(0.05/16{,}471)$).