International Journal of
Molecular Sciences

MDPI

Article

Genetic Architecture of Body Weight Variation and GWAS-Informed Genomic Selection in Formulated Diet-Adapted Mandarin Fish (*Siniperca chuatsi*)

Chenfan Geng [1,†], Hailiang Song [1,†], Yinglan Xia [1], Xin Zhang [1], Shuisheng Li [2], Shuang Liu [2], Jinmeng Teng [2], Yong Zhang [2] and Hongxia Hu [1,*]

[1] Fisheries Science Institute, Beijing Academy of Agriculture and Forestry Sciences & Beijing Key Laboratory of Fisheries Biotechnology, Beijing 100068, China; gengchenfan@baafs.net.cn (C.G.)
[2] State Key Laboratory of Biocontrol and School of Life Sciences, Southern Marine Science and Engineering Guangdong Laboratory (Zhuhai), Guangdong Provincial Key Laboratory of Aquatic Economic Animals and Guangdong Provincial Engineering Technology Research Center for Healthy Breeding of Important Economic Fish, Sun Yat-Sen University, Guangzhou 510275, China
* Correspondence: huhongxia@baafs.net.cn
† These authors contributed equally to this work.

Abstract

Large variation in body weight poses a major challenge to the development and industrialization of formulated-feed aquaculture in *Siniperca chuatsi*. Breeding varieties that combine formulated-diet adaptation, rapid growth, and uniform body weight is therefore important for improving production efficiency. However, the genetic architecture of body weight and optimal genomic selection (GS) strategies remain poorly understood. Here, whole-genome resequencing was performed on 522 formulated-diet-adapted individuals. Genome-wide association study (GWAS) identified 21 significant loci, with candidate genes associated with lipid metabolism, endocrine regulation, and neuroendocrine signaling. Genomic prediction was evaluated using genome-wide single-nucleotide polymorphisms (SNPs) and GWAS-prioritized SNP panels with Bayesian and machine-learning models. Prediction accuracy increased with marker number and generally plateaued at approximately 30,000 SNPs. Random forest using GWAS-prioritized markers achieved the highest accuracy, improving prediction accuracy by 4.8% over genome-wide genomic best linear unbiased prediction (GBLUP) while using fewer markers. These findings demonstrate the potential of GWAS-informed marker prioritization to reduce marker requirements and improve GS efficiency for body weight in formulated-diet-adapted *S. chuatsi*.

Keywords: *Siniperca chuatsi*; genomic prediction efficiency; genomic selection; SNP density; genetic improvement; formulated diet adaptation; aquaculture industry; China

Check for updates

Academic Editor: Salvatore Saccone

Received: 12 September 2026
Revised: 17 October 2026
Accepted: 19 October 2026
Published: 22 October 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

1. Introduction

Mandarin fish (*Siniperca chuatsi*) is an economically important freshwater aquaculture species in China, valued for its rapid growth, favorable flesh quality, and high consumer acceptance; reported per-unit economic returns are more than four times those of major freshwater fish species [1–3]. Production has continued to rise in response to market demand, reaching approximately 603,436 t in 2025, an 11.79% increase over 2024 [4]. However, conventional production still relies heavily on live forage fish, a practice that requires substantial land and labor inputs and increases the risks of pathogen transmission and disease

outbreaks, while also entailing considerable management costs [1,5,6]. Feed training offers a feasible route to formulated-diet acceptance in mandarin fish [1,7]. Nevertheless, after successful adaptation, growth performance remains highly variable among individuals, and body weight differs considerably among formulated-diet-adapted fish. The genetic basis of this variation is not well understood, constraining genetic improvement of growth under formulated-diet feeding. Clarifying the genetic basis of growth variation in formulated-diet-adapted mandarin fish and identifying individuals with high genetic potential for growth are therefore important for improving this trait through genetic selection.

Research on formulated-diet adaptation in mandarin fish has to date centered on comparing individuals that readily accept formulated diets with those that remain resistant to dietary transition, in order to elucidate the molecular basis of individual variation in feed-adaptation capacity [8–10]. Genomic studies have addressed the genetic basis of feeding-habit domestication. Early candidate-gene work identified polymorphisms in lipid metabolism and digestion genes, including *LPL* and *PEP*, as potential correlates of artificial-diet acceptance. Subsequent whole-genome resequencing identified 57 selective sweep regions and 149 candidate genes associated with formulated-diet domestication [8]. Epigenetic and epigenomic analyses have further implicated regulatory mechanisms. Increased DNA methylation of *TFIIF* and reduced H3K27me3 levels, accompanied by altered *EZH1* expression, were associated with formulated-diet acceptance [10]. Moreover, whole-genome bisulfite sequencing identified 5976 differentially methylated regions involving 3522 genes—including *PRKACBA*, *CREB5B*, *BDNF*, and *LPL*—with enrichment in cAMP signaling, lipid metabolism, and neural plasticity [8]. Transcriptomic and metabolomic profiling has revealed coordinated changes in retinol metabolism, glycerolipid metabolism, and unsaturated fatty acid biosynthesis; for example, expression of triacylglycerol lipase and stearoyl-CoA-related genes, together with corresponding metabolic changes, differed between formulated-diet-adapted and non-adapted individuals [10,11]. 16S rRNA profiling of the gut microbiome has shown marked differences in microbial diversity and community composition between fish fed formulated diets and those fed live prey, with *Lactococcus*, *Klebsiella*, and *Woeseia* enriched in larger formulated-diet-fed individuals and greater representation of microbial pathways related to glycan and amino acid metabolism [12]. More recently, integrated analyses of serum biochemical traits, gene expression, and gut microbiota in domestication-susceptible and domestication-resistant individuals have further characterized the physiological and molecular correlates of feed-adaptation capacity [13]. Together, these findings indicate that adaptation to formulated diets is a complex process involving coordinated changes across genomic, transcriptional, epigenetic, metabolic, and gut microbial processes. Nevertheless, the emphasis has remained largely on explaining why individuals differ in their capacity to adapt to formulated diets, rather than on the genetic basis of growth variation among individuals that have already adapted successfully.

Body weight is a complex quantitative trait governed by multiple genetic variants and environmental factors, and its genetic architecture is difficult to resolve through conventional approaches [14,15]. Chromosome-scale reference genomes and high-throughput genotyping technologies have enabled genome-wide analyses of economically important traits in mandarin fish [2,16,17]. Several GWAS have investigated growth-related traits in this species. An initial GWAS based on 5205 SLAF-seq SNPs detected 11 genome-wide significant and 37 suggestive loci across six growth traits and annotated 65 candidate genes in the surrounding regions [18]. A subsequent GWAS using 49,940 ddRAD-seq SNPs in the all-female "Dinggui I" strain identified six significant and 25 suggestive loci for five growth traits, with several candidate genes implicated in growth regulation [19]. These studies support a genetic basis for growth variation in mandarin fish and provide valuable genomic resources for genetic improvement. Existing growth-related GWAS, however, have focused mainly on conven-

tional breeding populations and have not specifically examined body weight variation in mandarin fish that have successfully adapted to formulated diets. The genetic determinants of growth performance under formulated-diet feeding thus remain inadequately characterized, and whether genomic information can effectively predict genetic merit for body weight in formulated-diet-adapted populations is largely unresolved.

Previous studies have mainly focused on the molecular mechanisms underlying formulated-diet adaptation, whereas the genetic basis of body weight variation among fish that have already adapted successfully remains poorly understood. This study therefore aimed to elucidate the genetic architecture of body weight variation in formulated-diet-adapted mandarin fish and to develop an efficient genomic prediction strategy for selective breeding. We first performed whole-genome sequencing-based GWAS to identify body weight-associated loci and candidate genes, and then incorporated GWAS-informed SNPs into genomic prediction. SNP panels with different marker densities and multiple prediction models were evaluated by cross-validation to identify a strategy with high predictive accuracy and reduced genotyping and computational requirements. By integrating GWAS with genomic prediction, this study provides insights into the genetic basis of growth variation in formulated-diet-adapted mandarin fish and establishes a practical genomic framework for improving body weight through selective breeding.

2. Results

2.1. Substantial Individual Variation in Body Weight Under Formulated Diet

A total of 522 individuals were measured for body weight, with a mean of 487.75 g (SD = 277.59 g) and a coefficient of variation of 56.91%. Individual weights ranged from 55 to 1461.2 g, a 26.6-fold difference between the lightest and heaviest individuals, indicating substantial phenotypic variation in this formulated-diet-adapted population. The narrow-sense genomic heritability (h^2) of body weight was estimated as 0.684 ± 0.081 using a GBLUP model based on the genomic relationship matrix. Table 1 summarizes these descriptive statistics and genetic parameters.

Table 1. Descriptive statistics and heritability estimate for body weight.

Category	Statistic	Value
Population	Number of fish	522
Body weight (g)	Mean ± SD	487.75 ± 277.59
	CV (%)	56.91
	Minimum	55
	Maximum	1461.2
Genetic parameter	Heritability (h^2) ± SE	0.684 ± 0.081

Note: SD means standard deviation; CV means coefficient of variation; SE means standard error.

2.2. Whole Genome Resequencing and Population Structure Analysis

Whole-genome sequencing was performed on 522 individuals. A total of 24.66 billion paired-end reads were generated, corresponding to an average of approximately 47.24 million paired-end reads per individual. The average mapping rate to the reference genome was 99.63%. Mean sequencing depth was 19.71× across all individuals, ranging from 13.65× to 29.17×. Initial variant calling identified 4,308,126 raw variants, including 2,926,398 SNPs. After stringent filtering, 2,299,387 high-quality SNPs were retained for downstream analyses. The genomic density distribution of these SNPs is shown in Figure 1a.

Principal component analysis (PCA) revealed genetic variation among the formulated-diet-accepted mandarin fish individuals, although no clear clustering pattern was observed (Figure 1b). To control for population stratification, GWAS models including the first 5 to 10 principal components (PCs) as covariates were evaluated. Among the evaluated models,

inclusion of the first seven PCs resulted in the least deviation of the genomic inflation factor from 1 (λ = 0.95; Supplementary Table S1), and these PCs were therefore retained in the final GWAS model. Genome-wide linkage disequilibrium (LD) decay was estimated using pairs of adjacent markers. As shown in Figure 1c, the genome-wide average squared correlation coefficient (r^2) was 0.167, and the corresponding physical distance was approximately 87.91 kb. Based on this decay pattern, a 90-kb window upstream and downstream of significantly associated SNPs was used to screen for candidate genes. The heatmap of the genomic relationship matrix (G matrix) for the 522 individuals (Figure 1d) showed prominent block-like aggregation consistent with the PCA results, reflecting the population structure and providing a suitable framework for genomic prediction.

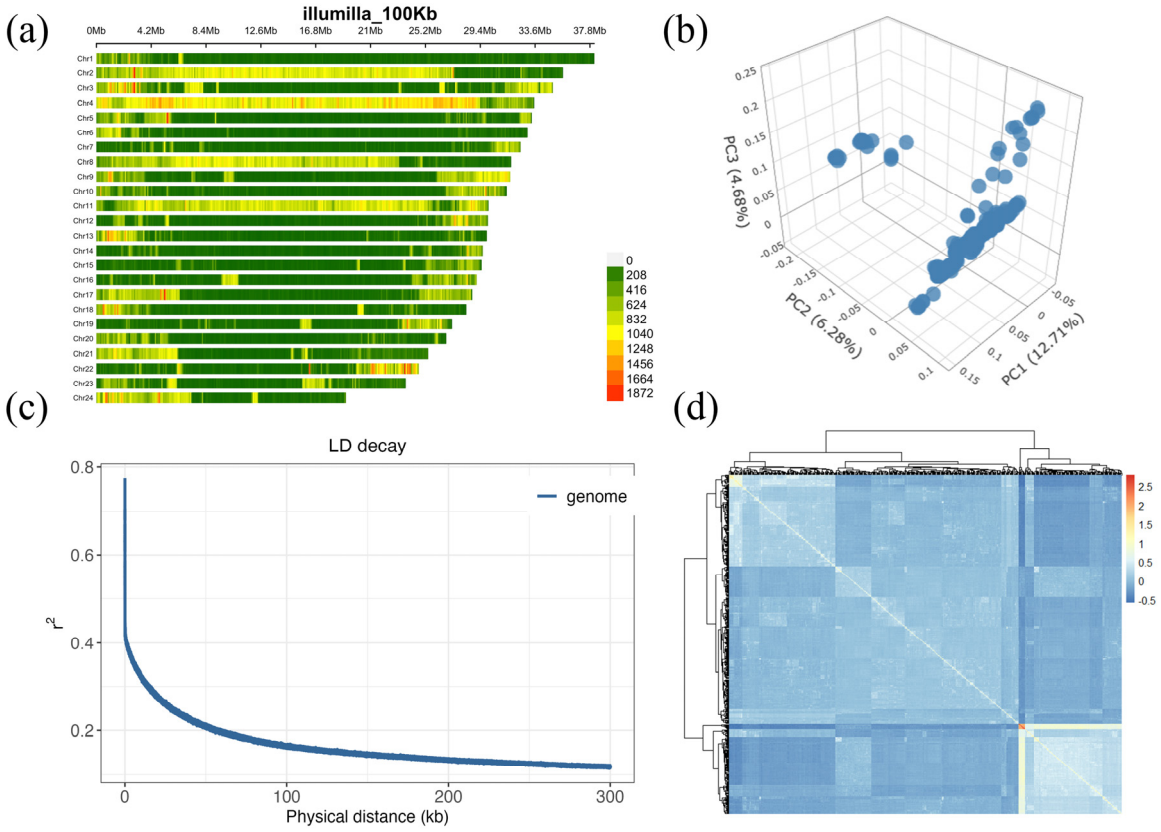


Figure 1. SNP distribution and population structure of mandarin fish. (**a**) Distribution of SNPs in 100 Kb windows across the genome; (**b**) principal component analyses for the first to the third dimensions of principal component (PC); (**c**) genome-wide LD decay; (**d**) heatmap of genomic relationship matrix.

2.3. Genome-Wide Association Analysis of Body Weight

A GWAS of body weight in formulated-diet-adapted mandarin fish was performed using whole-genome resequencing data. Using Bonferroni correction, 21 genome-wide significant SNPs were identified (Figure 2a; Supplementary Table S2). These SNPs were distributed on two chromosomes, with 2 on chromosome 4 and 19 on chromosome 6; chromosome 6 harbored the majority of associated SNPs. The QQ plot showed no apparent deviation from the expected null distribution (Figure 2b). Within ±90 kb of the significant SNPs, 132 protein-coding genes were annotated and considered candidate genes for body weight variation (Supplementary Table S3).

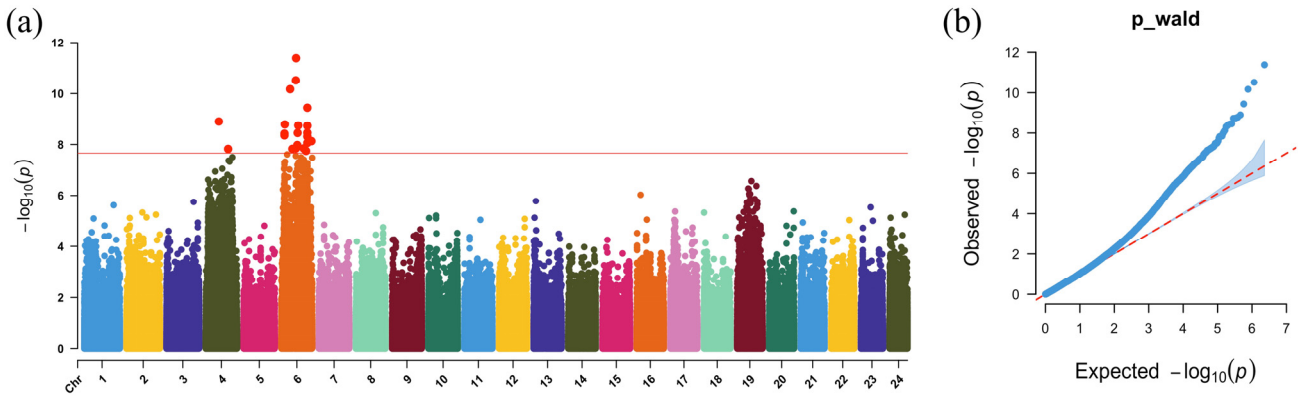


Figure 2. GWAS identified SNPs significantly associated with body weight trait in *S. chuatsi*. (**a**) Manhattan plots for body weight trait. SNPs from different chromosomes are displayed in different colors, with each color representing SNPs located on a different chromosome. Red dots indicate significant SNPs that exceed the Bonferroni-corrected significance threshold. The red line represents the Bonferroni-corrected significance threshold. (**b**) Quantile-quantile plots of GWAS for body weight trait. GWAS (genome-wide association study).

2.4. Gene Annotation and Functional Enrichment Analysis of Candidate Genes

Enrichment analysis of the 132 candidate genes identified 87 GO terms and 26 KEGG pathways showing nominal enrichment based on unadjusted p values ($p < 0.05$; Supplementary Tables S4 and S5). These nominally enriched terms and pathways could be broadly categorized into two functional themes: lipid metabolic and endocrine pathways, and processes related to neuroendocrine regulation of feeding behavior and energy homeostasis.

2.4.1. Enrichment of Nutrient Utilization and Growth Associated Metabolic Functions and Pathways

Enrichment analysis identified GO terms and KEGG pathways associated with cellular processes related to growth, lipid metabolism, and nutrient utilization (Figure 3). Among these, GO terms involved in lipid digestion, transport, and deposition were enriched, including plasma lipoprotein particle organization and remodeling, positive regulation of lipid storage, lipase activity, and phospholipase A_2 activity. These terms were represented mainly by *LCAT*, *LPL*, *PRKACBA*, *PLA2G15*, and *LOC122877981*. KEGG analysis further identified several metabolic and endocrine-related pathways, including insulin secretion, cholesterol metabolism, thyroid hormone synthesis, and regulation of lipolysis in adipocytes. These findings suggest that body weight-associated candidate genes may contribute to post-feeding nutrient utilization, growth-related metabolic regulation, lipid metabolism, and endocrine metabolic response.

2.4.2. Enrichment of Neuroendocrine Signaling and Feeding-Associated Functions and Pathways

A second set of nominally enriched functions was associated with neuroendocrine signaling and feeding-related regulation. GO enrichment identified terms including cAMP-mediated signaling, inositol phosphate biosynthetic process, cyclic nucleotide-mediated signaling, positive regulation of phospholipase C activity, and multicellular organismal response to stress (Figure 4a). These terms involved candidate genes such as *ADCYAP1R1B*, *PRKACBA*, *CDC34A*, *PEX5LA*, *LOC122877981*, *LOC122878068*, *LOC122877219*, and *USP46* (Figure 4b). KEGG pathways related to neural signaling and feeding regulation included circadian entrainment, serotonergic synapse, retrograde endocannabinoid signaling, GABAergic synapse, dopaminergic synapse, and the apelin signaling pathway. Several genes,

including *PRKACBA*, *LOC122877870*, *LOC122877889*, *LOC122877981*, and *LOC122877219*, were involved in multiple pathways (Figure 5).



Figure 3. GO and KEGG enrichment analysis of body weight-associated candidate genes related to nutrient utilization and growth-associated metabolic functions. The size of the dots represents the number of enriched genes (Count), and the color of the dots indicates the enrichment category (GO: BP, CC, MF, and KEGG pathway). BP: biological process; CC: cellular component; MF: molecular function; KEGG pathway: Kyoto Encyclopedia of Genes and Genomes pathway.

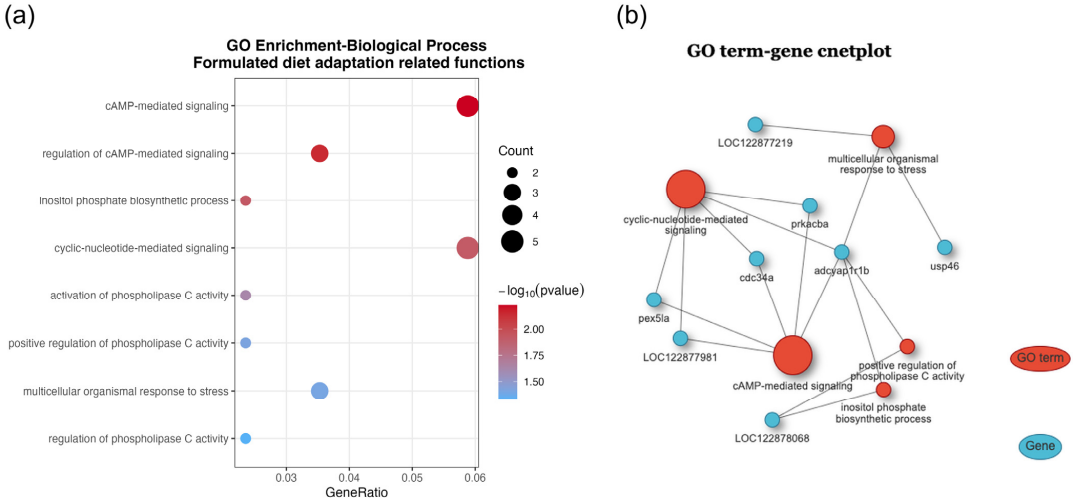


Figure 4. GO enrichment analyses of formulated diet adaptation-related neuroendocrine signaling and feeding-associated functions. (**a**) Dotplot of GO enrichment analysis ($p < 0.05$). The x-axis represents the GeneRatio, the size of the dots represents the number of enriched genes (Count), and the color gradient from blue to red indicates the significance level ($-\log_{10}(p\text{-value})$). (**b**) Cnetplot of GO terms and genes. The red nodes represent GO terms, the cyan nodes represent genes, and the edges indicate the connections between them.

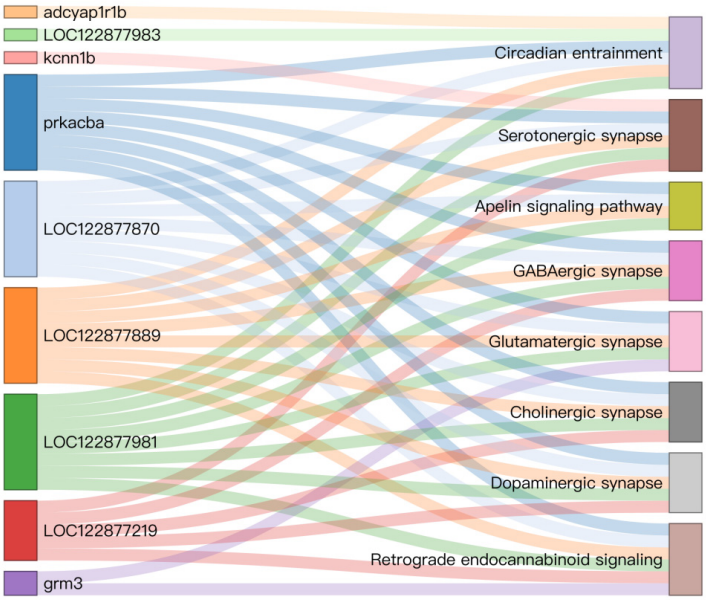


Figure 5. KEGG enrichment analyses of neuroendocrine signaling and feeding-associated pathways.

2.5. Effect of GWAS-Selected Marker Density on Genomic Prediction Performance for Body Weight

Prediction accuracy for body weight depended on both GWAS-selected marker density and prediction model (Figure 6). At 1 K markers, median prediction accuracy for most models was below the whole-genome sequencing (WGS)-based genomic best linear unbiased prediction (GBLUP) benchmark. Accuracy generally increased with marker density. Random forest (RF) was the first model to exceed the benchmark, at 3 K markers, followed by extreme gradient boosting (XGB). At 10 K markers, several models reached or exceeded the benchmark. At 30 K markers, all Bayesian and machine-learning models exceeded the benchmark. At 50 K markers, most models performed comparably to or better than the benchmark. Across all marker densities, RF achieved the highest prediction accuracy among the evaluated methods.

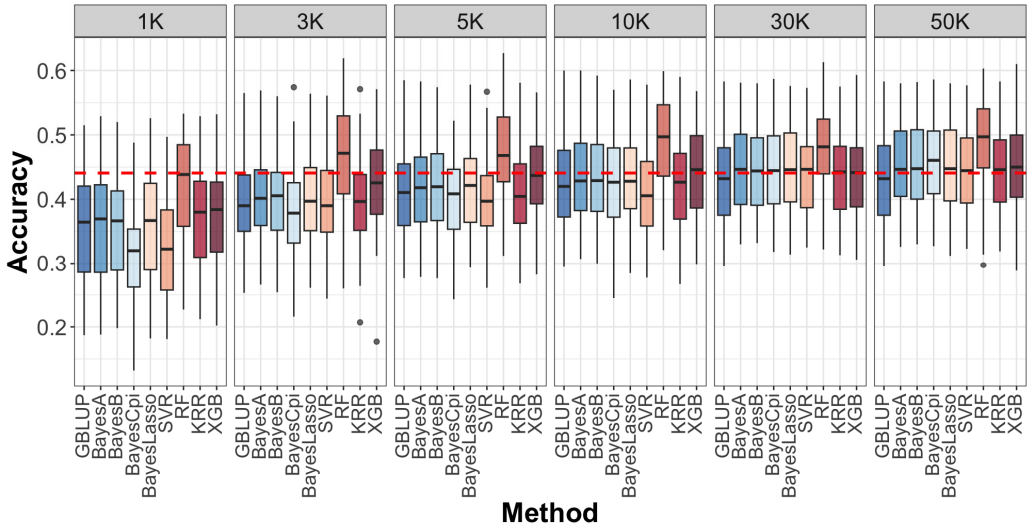


Figure 6. Accuracy of genomic prediction for body weight using GWAS-selected marker sets (1 K–50 K). Linear models included GBLUP, BayesA, BayesB, BayesCπ, and BayesLasso; machine-learning methods included SVR, RF, KRR, and XGB. The red dashed line indicates the prediction accuracy of the WGS-based GBLUP benchmark. Abbreviations: GBLUP, genomic best linear unbiased prediction; GWAS, genome-wide association study; KRR, kernel ridge regression; RF, random forest; SVR, support vector regression; WGS, whole-genome sequencing; XGB, extreme gradient boosting.

The calibration slopes of genomic prediction for body weight are shown in Figure 7. At 1 K markers, most models had median slopes below the WGS-based GBLUP benchmark. With increasing marker density, the slopes generally increased and approached the benchmark. RF reached or exceeded the benchmark at 3 K markers, followed by extreme gradient boosting (XGB) and kernel ridge regression (KRR) at higher marker densities. At 30 K and 50 K markers, RF, KRR, and XGB had median slopes comparable to or slightly above the benchmark, whereas GBLUP and the Bayesian models remained below the benchmark. Among all methods, RF showed the highest median calibration slope across marker densities, although its slope distribution was wider and included several outliers. Overall, calibration slopes increased with GWAS-selected marker density, and RF consistently showed the highest median slope across marker densities.

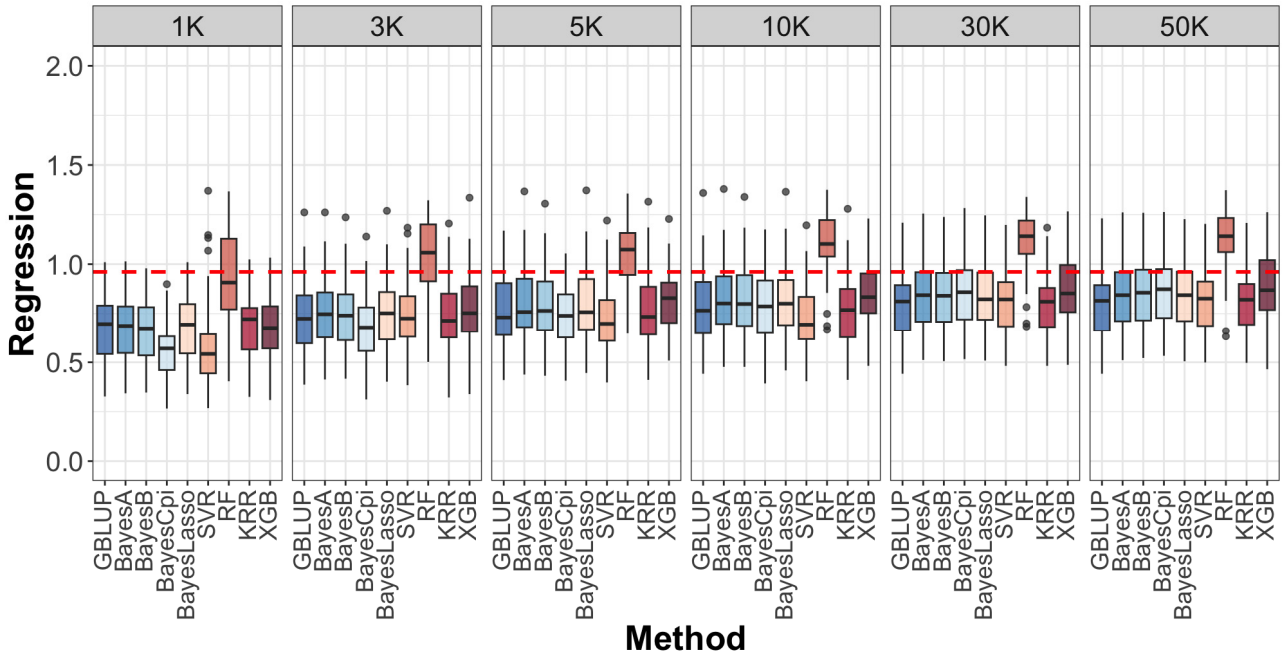


Figure 7. Regression coefficients (calibration slopes) of genomic prediction for body weight using GWAS-selected marker sets (1 K–50 K markers). Observed body weight was regressed on predicted genomic estimated breeding values (GEBVs). The red dashed line indicates the WGS-based GBLUP benchmark. Models and abbreviations are as in Figure 6.

3. Discussion

3.1. Genetic Architecture Underlying Body Weight Variation Following Formulated Diet Adaptation

We characterized the genetic architecture of body weight variation in formulated-diet-adapted *S. chuatsi*. Despite uniform diet adaptation and subsequent pond rearing, body weight varied widely among individuals, and the estimated narrow-sense genomic heritability was 0.684 ± 0.081, indicating a substantial genetic contribution. GWAS detected genome-wide significant associations concentrated on chromosome 6, with a smaller number on chromosome 4, consistent with polygenic control and a prominent hotspot on chromosome 6. Functional annotation of the candidate genes implicated nutrient utilization and growth-associated metabolism, as well as neuroendocrine regulation of appetite and energy homeostasis. Previous studies on formulated-diet adaptation in mandarin fish have mainly focused on diet acceptance and epigenetic or expression changes [8,12,13,20,21]. The present study extends these findings by linking genomic variation to growth performance after successful adaptation.

3.1.1. Lipid Metabolic and Endocrine Pathways Underlying Body Weight Variation

The present GWAS identified candidate genes and enriched pathways associated with body weight variation in formulated-diet-adapted mandarin fish, with prominent signals related to lipid metabolism and endocrine regulation. The co-occurrence of *LPL*, *LCAT*, and *PLA2G15* suggests that growth differences may involve variation in lipid acquisition, transport, remodeling, and intracellular utilization after adaptation to formulated diets.

Among these candidates, *LPL* is of particular interest. As a key enzyme that hydrolyzes triglyceride-rich lipoproteins and facilitates fatty acid uptake by peripheral tissues, *LPL* regulates lipid availability for energy metabolism and tissue growth [22]. In mandarin fish, *LPL* polymorphisms, methylation patterns, and hepatic expression differences have been associated with formulated-diet acceptance [8,20,22]. The *LPL* variants associated with body weight here extend these findings by suggesting that lipid utilization may influence not only the capacity to accept formulated diets but also subsequent growth performance among adapted individuals. In addition to *LPL*, *LCAT* and *PLA2G15* were identified, indicating that multiple stages of lipid metabolism may contribute to growth variation. *LCAT* participates in cholesterol esterification and lipoprotein remodeling, whereas *PLA2G15* is involved in lysosomal phospholipid turnover and intracellular lipid homeostasis [23,24]. Although direct evidence linking these genes to fish growth traits remains limited, their functions, together with those of *LPL*, suggest that coordinated regulation of lipid processing, rather than a single metabolic pathway, may underlie body weight differences under formulated-diet conditions.

Beyond lipid metabolism, enrichment of insulin secretion, thyroid hormone synthesis, and lipolysis-related pathways points to a possible role of endocrine regulation in metabolic variation. In particular, *PRKACBA* and *LOC122877981* may participate in cAMP-mediated signaling, which integrates hormonal and nutritional cues to regulate downstream metabolic processes. Previous epigenetic studies have also implicated cAMP-related regulation, including altered methylation of *PRKACBA*, in mandarin fish differing in formulated-diet acceptance [8]. Although the present GWAS associations do not establish causal effects and require functional validation, they support a genetic framework in which lipid metabolic capacity and endocrine regulation jointly contribute to body weight variation in formulated-diet-adapted mandarin fish. In contrast to previous studies that focused primarily on the genetic basis of formulated-diet acceptance, the present findings suggest that variation in metabolic efficiency after successful adaptation may be an important determinant of growth performance.

3.1.2. Neuroendocrine Regulation of Feeding Behavior and Energy Homeostasis

Beyond lipid metabolism, the candidate genes identified in the present study were enriched in neuroendocrine-related processes, including cAMP signaling, regulation of phospholipase activity, stress responses, and multiple neurotransmitter synaptic pathways. Several genes, including *ADCYAP1R1B*, *PRKACBA*, *CDC34A*, *PEX5LA*, *LOC122877981*, and *LOC122878068*, contributed to these enriched categories, suggesting that variation in neuroendocrine regulation may represent another mechanism underlying body weight differences among formulated-diet-adapted individuals.

Among these genes, *ADCYAP1R1B* showed particularly strong functional relevance, as it was repeatedly detected across multiple enriched GO categories. This gene encodes a receptor for pituitary adenylate cyclase-activating polypeptide (PACAP), a neuropeptide involved in the regulation of feeding-related endocrine responses in teleosts [25]. Previous studies have suggested that PACAP signaling can influence appetite regulation through interactions with hypothalamic neuroendocrine pathways, including corticotropin-releasing hormone (CRH)-mediated signaling [25–27]. In mandarin fish, previous genomic studies of

feeding habit domestication mainly identified selection signals and regulatory changes associated with the transition from live prey to formulated diets, particularly involving sensory perception, neural function, and metabolic regulation [8,20]. In contrast, the present study identified variation in *ADCYAP1R1B* and related neuroendocrine genes among individuals that had already acquired formulated diet acceptance, suggesting that these pathways may contribute not to the initial establishment of feeding behavior but to differences in feeding regulation, physiological state, and growth potential after adaptation.

The enrichment of stress-related biological processes further indicates that neuroendocrine regulation may influence growth indirectly by affecting feeding stability and energy allocation. Previous studies have shown that formulated diet domestication can induce changes in oxidative stress responses and antioxidant capacity in mandarin fish [28]. Therefore, genetic variation affecting stress responsiveness may influence the ability of individuals to maintain stable feeding activity and allocate energy toward growth rather than physiological stress responses.

KEGG enrichment analysis further revealed enrichment of multiple neurotransmission-related pathways, including serotonergic, GABAergic, glutamatergic, cholinergic, and dopaminergic synapses, as well as the apelin signaling pathway. These pathways are known to participate in the neural regulation of feeding and energy homeostasis. Previous studies have shown that serotonergic, GABAergic, and endocannabinoid systems contribute to central control of food intake [29], while dopaminergic circuits can modulate feeding motivation and satiety [30]. In teleosts, apelin signaling has also been implicated in appetite and growth regulation, with Pyr-apelin-13 altering the expression of appetite- and growth-related genes in common carp [31] and apelin functioning as an orexigenic factor in goldfish [32]. Similarly, the endocannabinoid system has been shown to promote food intake in rainbow trout and goldfish [33,34]. In the present study, several candidate genes, including *PRKACBA*, *LOC122877870*, *LOC122877889*, *LOC122877981*, and *LOC122877219*, were involved in or shared across these feeding-related pathways, suggesting that body weight variation among formulated-diet-adapted mandarin fish may be influenced by coordinated regulation of neuroendocrine networks controlling appetite, feeding behavior, and energy allocation. These findings extend previous studies on formulated-diet domestication by indicating that neuroendocrine pathways may continue to contribute to phenotypic variation after feeding adaptation has been established. Whereas earlier studies focused primarily on the genetic basis of whether mandarin fish can accept formulated diets, the present results suggest that individual differences in appetite-related regulation, stress responsiveness, and energy homeostasis may further influence growth performance under formulated-diet conditions. Functional validation of these candidate genes will be necessary to clarify their roles in feeding efficiency and body weight variation.

3.2. Effects of SNP Density on Genomic Prediction

Prediction accuracy increased with SNP density under GWAS-informed marker selection, with limited improvement beyond approximately 30 K SNPs. Similar trends have been reported in other aquaculture species, where GWAS-informed SNP subsets achieved prediction accuracy comparable to or higher than that of full WGS marker sets for growth and production traits in sturgeon, Atlantic salmon, common carp, and rainbow trout [35,36]. At this density and above, GWAS-prioritized panels consistently achieved high prediction accuracy for body weight, indicating that a moderate-density panel may be sufficient for genomic prediction in this population. Genomic relationship matrix analyses supported this pattern: the correlation between the WGS-based G matrix and SNP subset-derived matrices increased from 0.74 at 1 K SNPs to 0.88 at 50 K SNPs, whereas matrices based on

10 K–50 K SNPs were highly correlated (r = 0.99–1.00), indicating that genomic relationship estimates stabilized at moderate-to-high marker densities (Supplementary Figure S1).

Increasing marker number alone did not necessarily improve prediction. Although WGS provides broad genome coverage, some variants may contribute limited additional information because of redundancy or weak associations with the target trait. Low- or medium-density panels combined with imputation can maintain high prediction accuracy when training and validation populations are closely related [37,38]. In the present population, a panel of approximately 30 K informative SNPs therefore appears to balance prediction accuracy and genotyping efficiency for body weight improvement in formulated-diet-adapted mandarin fish. Reducing marker density while preserving predictive performance would make genomic selection more cost-effective in future breeding programs.

3.3. Comparison of Linear and Machine Learning Models on Genomic Prediction

Genomic prediction performance for body weight in formulated-diet-adapted mandarin fish depended on both SNP density and model choice. Among linear models, BayesA, BayesB, and BayesLASSO generally yielded slightly higher prediction accuracies than GBLUP, whereas BayesCπ showed lower accuracy at low SNP densities but improved as marker density increased, reaching the highest median accuracy among Bayesian models at 50 K SNPs. Overall, differences among Bayesian models and GBLUP were modest, consistent with the polygenic architecture of body weight, in which many loci of small effect collectively contribute to phenotypic variation.

Machine-learning models differed more markedly in performance. Random forest (RF) consistently achieved the highest or near-highest accuracy across SNP densities and outperformed GBLUP, Bayesian models, Support vector regression (SVR), KRR, and XGB in most scenarios. This advantage may reflect the ability of RF to capture nonlinear relationships and interactions among markers without predefined assumptions about marker effects [39]. However, RF also showed regression coefficients above the reference value, indicating potential prediction bias despite its superior accuracy. Both accuracy and calibration therefore need to be considered when evaluating genomic prediction models.

Model performance also varies with trait architecture, species, and population characteristics. In sturgeon, prioritized SNP selection from whole-genome sequencing improved prediction accuracy, and machine-learning models performed competitively relative to conventional linear approaches [36], supporting the present finding that biologically informed marker selection combined with suitable prediction models can enhance genomic prediction efficiency. In Holstein heifers, machine-learning and conventional models using low-coverage sequencing data achieved comparable performance for fertility traits, highlighting the potential of cost-effective strategies [35]. In rainbow trout, BayesC achieved relatively high accuracy for resistance to *Piscirickettsia salmonis*, and a limited number of SNPs sufficed to match higher-density panels [40]. In eastern oyster, BayesB outperformed Bayesian ridge regression and RKHS for low-salinity tolerance, suggesting that models allowing heterogeneous marker effects are advantageous when loci have unequal effects [41]. These findings are consistent with the present results, where Bayesian models performed stably but their relative advantage depended on SNP density.

Machine-learning approaches have also shown promise for complex aquaculture traits. In olive flounder (*Paralichthys olivaceus*), RF, Elastic Net, and BayesB achieved higher accuracies than GBLUP for growth traits [42], supporting the superior performance of RF observed here. Similarly, artificial intelligence-based algorithms improved genomic prediction for disease resistance in striped catfish compared with conventional methods [43], and machine-learning models showed advantages for growth-trait prediction in Pacific

white shrimp [44]. However, the optimal model differed across studies, indicating that prediction performance depends on the interaction between model assumptions, genetic architecture, marker density, and population structure. Combining GWAS-informed SNP selection with an appropriate prediction model therefore provides an efficient strategy for genomic selection of body weight in formulated-diet-adapted mandarin fish. Among the approaches evaluated, RF showed the highest predictive potential, whereas Bayesian models offered relatively stable performance across SNP densities. These results provide a practical basis for optimizing genomic prediction and reducing genotyping costs in future breeding programs.

3.4. Influence of GWAS-Based SNP Prioritization on Genomic Prediction Accuracy

The present findings indicate that GWAS-based SNP prioritization can improve genomic prediction accuracy in this population, consistent with previous studies. In sturgeon, GWAS-prioritized SNPs increased accuracy by 3.4% over the full WGS dataset at moderate marker densities (≥10 K) [36]. In *Siniperca scherzeri*, 500 GWAS-selected SNPs achieved a prediction ability of 0.70 for body length and total length [45]. A GBLUP variant incorporating BLINK-derived GWAS information (GAGBLUP) improved accuracy by more than 20% over conventional GBLUP using all genome-wide markers [46]. These results support GWAS-assisted marker selection as an efficient strategy for genomic prediction in specific populations and traits.

Model evaluation should also consider error-based metrics. In the present study, GBLUP consistently showed the highest median mean squared error (MSE) and mean absolute error (MAE) across SNP densities, whereas the evaluated Bayesian and machine-learning models had substantially lower and comparable error distributions (Supplementary Figures S2 and S3). This indicates that, although GBLUP served as a useful benchmark, Bayesian and machine-learning approaches achieved lower prediction errors for body weight in this population. Because RF showed high accuracy but also regression bias, model selection should jointly consider accuracy, calibration, and error-based metrics.

3.5. Limitations and Future Perspectives

Several limitations should be considered when interpreting these findings. Although GWAS identified significant SNPs and candidate genes associated with body weight variation, their biological functions and potential regulatory mechanisms remain to be experimentally validated. Future studies should focus on functional characterization of these candidate genes to better understand the molecular mechanisms underlying growth variation in formulated-diet-adapted mandarin fish. Furthermore, multi-generational selection experiments are needed to evaluate the long-term effectiveness and practical applicability of GWAS-informed SNP panels in breeding programs. Integrating genomic information with transcriptomic, epigenomic, or metabolomic data may further improve our understanding of growth regulation and enhance genomic prediction accuracy.

4. Materials and Methods

4.1. Population and Phenotype

All animal procedures complied with animal care laws and guidelines and were approved by the Ethics Committee of the Beijing Academy of Agriculture and Forestry Sciences (BAAFS-KJLL-20250509; 9 May 2025). Healthy juvenile all-female *S. chuatsi* were obtained from Guangdong Liangshi Aquatic Seed Industry Co., Ltd. (Foshan, China) and transported to the Sanshui Research and Experimental Base of the Pearl River Fisheries Research Institute, Chinese Academy of Fishery Sciences (Foshan, Guangdong, China). Fish

were initially fed live fry of mud carp (*Cirrhinus molitorella*). At approximately 4 cm body length, they underwent a 25-day gradual transition from live prey to a formulated diet, following the procedure described previously [47]. During adaptation, individuals that fed actively, swam vigorously, and had well-filled abdominal cavities were retained, whereas those with poor feed acceptance or low vitality were removed. Representative images of individuals with high and low feed-adaptation performance are shown in Supplementary Figure S4. The retained individuals constituted the formulated-diet-adapted population.

The adapted fish were subsequently reared in a single earthen pond (~0.67 ha) on the same formulated diet until 1 year of age. Water quality was monitored daily. Temperature, pH, and dissolved oxygen were measured using a YSI multiparameter water quality meter (YSI Inc., Yellow Springs, OH, USA), and ammonia nitrogen and nitrite nitrogen were determined using commercial test kits according to the manufacturer's instructions. Over the one-year rearing period, water quality was maintained as follows: temperature averaged 26.3 ± 6.5 °C (range, 20–32 °C), dissolved oxygen > 6 mg/L, pH averaged 7.9 ± 0.23, ammonia nitrogen < 0.1 mg/L, and nitrite nitrogen < 0.05 mg/L. Water exchanges were performed with temperature-equilibrated water. No obvious clinical signs or abnormal behaviors were observed during the rearing period.

A total of 522 one-year-old fish were randomly selected for phenotypic measurement and genomic analysis. Fish were anesthetized with eugenol, weighed using an electronic balance (±0.01 g), and caudal fin tissues were collected and preserved in 95% ethanol for whole-genome resequencing. Descriptive statistics for body weight (BW), including minimum, maximum, mean, standard deviation (SD), and coefficient of variation (CV), were calculated.

4.2. Whole-Genome Resequencing and Variant Calling

Genomic DNA was extracted from caudal fin tissue using the DNeasy Blood & Tissue Kit (Qiagen, Hilden, Germany) [48]. DNA concentration and integrity were assessed using a NanoDrop microvolume spectrophotometer (Thermo Fisher Scientific, Waltham, MA, USA) and 1.0% agarose gel electrophoresis, respectively. Paired-end libraries with 150-bp reads were prepared and sequenced on the DNBSEQ T7 platform (MGI Tech Co., Ltd., Shenzhen, China) by BGI Genomics Co., Ltd. (Shenzhen, China), with an expected average sequencing depth of approximately 20× per individual.

Raw reads were processed using fastp (v0.23.2) [49] to remove adapter sequences, low-quality bases, and reads containing >10% N bases. Clean reads were aligned to the *Siniperca chuatsi* reference genome (assembly ASM2008510v1) using BWA-MEM (v0.7.17) [50]. Aligned reads were sorted using SAMtools (v1.19) [51], and PCR duplicates were removed using sambamba (v1.0.1) [52]. Variants were called using GATK HaplotypeCaller (v4.1.3.0) following the best-practice workflow, and the resulting gVCF files were combined using CombineGVCFs. Variant quality filtering was performed using VariantFiltration with the following criteria: QUAL < 30.0, QD < 2.0, FS > 60.0, MQ < 40.0, MQRankSum < −12.5, and ReadPosRankSum < −8.0. Missing genotypes were imputed using Beagle 4.1 [53]. Final SNP quality control was performed using PLINK (v1.90) [54], retaining biallelic SNPs with a call rate >90%, minor allele frequency >0.01, and Hardy–Weinberg equilibrium $p > 1 \times 10^{-6}$.

4.3. Population Structure and Linkage Disequilibrium Analysis

Population structure was evaluated by PCA based on the genomic relationship matrix constructed from genome-wide SNPs using GCTA (v1.25.3) [55]. The first three principal components were used to visualize genetic relationships among individuals. To control for potential population stratification in the GWAS, the first 5 to 10 principal components were evaluated as covariates, and the corresponding genomic inflation factor (λ) was calculated

for each model (Supplementary Table S1). A λ value closer to 1 indicates better control of potential inflation of association statistics and was considered when selecting the optimal number of PCs [56].

Genome-wide linkage disequilibrium was quantified using the squared correlation coefficient (r^2) with PopLDdecay (v3.42) [57]. The physical distance corresponding to the genome-wide average r^2 was used to define the LD decay distance and the candidate-gene search window.

4.4. Genome-Wide Association Analysis

Genome-wide association analysis was performed using the linear mixed model (LMM) implemented in GEMMA (v0.98.5) [58] to identify single-nucleotide polymorphisms (SNPs) associated with body weight. A genetic relationship matrix (GRM) was constructed from all SNPs to account for genetic relatedness among individuals. The narrow-sense genomic heritability (h^2) of body weight was estimated using a GBLUP model based on the GRM as $h^2 = \frac{\sigma_g^2}{\sigma_g^2 + \sigma_e^2}$, where σ_g^2 and σ_g^2 represent additive genetic and residual variances, respectively. All three tests (Wald, likelihood ratio, and score) were performed using the -lmm 4 option. Genome-wide significance was determined using the Bonferroni-corrected threshold ($p < 0.05/N$), where N is the number of SNPs tested. Manhattan and quantile–quantile (QQ) plots were generated using the CMplot package [59] in R. Based on the LD decay distance determined in Section 4.3, candidate genes were annotated within ±90 kb of significant SNPs. For genomic prediction, GWAS-based SNP selection was performed independently within the training population of each cross-validation fold, excluding the corresponding validation population from GWAS and SNP prioritization, to prevent information leakage. The selected SNP panels were then used for prediction in the corresponding validation population.

4.5. Gene Annotation and Functional Enrichment Analysis

Protein-coding genes of *S. chuatsi* (assembly ASM2008510v1) were functionally annotated using eggNOG-mapper v2.1.15 [60]. Candidate genes identified by GWAS were subjected to Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analyses using the clusterProfiler R package (v4.3.3) [61]. Multiple-testing correction was performed using the Benjamini–Hochberg (BH) method [62]. Enriched terms and pathways were evaluated based on both nominal p values and BH-adjusted p values, with $p < 0.05$ and adjusted $p < 0.05$ considered nominal and statistically significant enrichment thresholds, respectively.

4.6. GWAS-Informed SNP Panel Construction

To determine whether GWAS-informed marker reduction could preserve predictive information while decreasing marker density, SNPs were ranked by ascending GWAS p-values. Six panels containing the top 1 K, 3 K, 5 K, 10 K, 30 K, and 50 K SNPs were then constructed [36]. These reduced panels were compared with the full genome-wide SNP set to assess the effect of marker density on genomic prediction. Because the full SNP set was computationally prohibitive for Bayesian and machine-learning models, GBLUP was used as the benchmark for the full SNP set, whereas the six GWAS-informed panels were evaluated using GBLUP [63], Bayesian models, and machine-learning algorithms.

4.7. Genomic Prediction Models

Genomic prediction of BW was performed using GBLUP, four Bayesian alphabet models (BayesA and BayesB [64], BayesCπ [65], and Bayesian LASSO [66]), and four machine-learning algorithms: SVR [67], RF [68], KRR [69], and XGBoost [70].

GBLUP was implemented in DMU v6 [71]. Genomic relationship matrices were constructed from the full WGS-derived SNP set and from training-fold-specific GWAS-selected SNP panels. Variance components were estimated by restricted maximum likelihood using the average information algorithm, and GEBVs were obtained using the DMUAI procedure.

Bayesian models were implemented using the BGLR R package v1.1.4 [72]. These models differ in their assumptions regarding the distribution and magnitude of SNP effects, allowing evaluation of alternative genetic architectures. For each Bayesian model, 18,000 Markov chain Monte Carlo (MCMC) iterations were run, with the first 3000 discarded as burn-in. GEBVs were calculated as the posterior means of genetic values from the retained iterations.

SVR, RF, KRR, and XGBoost were implemented in Python v3.8.3 using scikit-learn v1.3.0 [73–75] and XGBoost v1.7.6 [70]. Hyperparameters were optimized by grid search with an inner cross-validation loop within each training fold, using the same training-fold-specific marker panel. The validation fold was not used for marker selection, hyperparameter tuning, or model training. Because the full WGS marker set was computationally demanding for Bayesian and machine-learning models, these models were evaluated using the six GWAS-selected marker densities (1 K, 3 K, 5 K, 10 K, 30 K, and 50 K). GBLUP was additionally evaluated using the full WGS-derived SNP set as the benchmark.

4.8. Benchmarking of Genomic Prediction Accuracy

All models were evaluated using repeated five-fold cross-validation [76]. The 522 individuals were randomly divided into five approximately equal-sized folds; four folds were used for training and the remaining fold for validation. Each fold served once as the validation set. The entire five-fold procedure was repeated 20 times using independent random partitions [77]. Within each repetition, identical fold assignments were used across all models and marker densities to ensure a fair comparison, and random seeds were fixed to ensure reproducibility. Marker panels were constructed as described in Section 4.6. In brief, GWAS was performed independently within the training population of each cross-validation fold, and the top SNPs were used to construct the 1 K, 3 K, 5 K, 10 K, 30 K, and 50 K panels; the validation population was excluded from GWAS and SNP prioritization. Predictions for each validation fold were generated using the model and marker panel trained on the corresponding training folds.

Prediction performance was assessed using the Pearson correlation coefficient between observed BW and predicted GEBVs [78]. Prediction bias was assessed by regressing observed BW on predicted GEBVs; the regression slope (calibration slope) was reported, with a slope of 1 indicating no systematic bias. MSE and MAE were additionally calculated to assess prediction error and overall model performance [77].

5. Conclusions

This study represents the first application of genomic selection for body weight prediction in formulated-diet-adapted *S. chuatsi*. Body weight exhibited substantial phenotypic variation and high genomic heritability, indicating considerable potential for genetic improvement through selection. By integrating GWAS with genomic prediction, we demonstrated that GWAS-prioritized SNP panels, particularly those containing approximately 30 K markers, achieved prediction performance comparable to or better than the whole-genome marker set, with random forest showing the highest accuracy among evaluated models. The identified candidate genes and pathways provide insights into the genetic basis of body weight variation after formulated-diet adaptation and support the application of GWAS-informed genomic selection in aquaculture breeding.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/ijms27219433/s1.

Author Contributions: Conceptualization, C.G., H.S. and H.H.; methodology, C.G., H.S. and Y.X.; software, C.G.; formal analysis, C.G. and H.S.; investigation, X.Z., S.L. (Shuisheng Li), S.L. (Shuang Liu) and J.T.; data curation, X.Z. and S.L. (Shuang Liu); writing—original draft preparation, C.G. and Y.X.; writing—review and editing, H.S., Y.Z. and H.H.; visualization, C.G.; supervision, Y.Z. and H.H.; project administration, H.H.; funding acquisition, H.H. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the Scientific and Technological Innovation Capacity Building Project of Beijing Academy of Agricultural and Forestry Sciences, China (Grant No. KJCX20261001). This work was also supported by the Scientific and Technological Innovation Capacity Building Project of Beijing Academy of Agricultural and Forestry Sciences, China (Grant Nos. KJCX20230216, KJCX20251009 and KJCX20261411); the China Agriculture Research System, funded by the Ministry of Finance and the Ministry of Agriculture and Rural Affairs of China (Grant No. CARS-46); the Excellent Young Scientist Program of Beijing Academy of Agricultural and Forestry Sciences (Grant No. YKPY2025004); the Beijing Key Laboratory of Fisheries Biotechnology (PT2025-20); and the National Key R&D Program of China (Grant Nos. 2024YFD2401500 and 2024YFD2401505).

Institutional Review Board Statement: The animal study protocol was approved by the Ethics Committee of Beijing Academy of Agriculture and Forestry Sciences (protocol code BAAFS-KJLL-20250909 and date of approval 9 September 2025).

Informed Consent Statement: Not applicable.

Data Availability Statement: The datasets used during the current study are available from the corresponding author on reasonable request.

Acknowledgments: The authors thank all members of the research group for their assistance and constructive discussions during this study.

Conflicts of Interest: All authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

GBLUP	Genomic best linear unbiased prediction
GEBV	Genomic estimated breeding value
GRM	Genetic relationship matrix
GS	Genomic selection
GWAS	Genome-wide association study
KRR	Kernel ridge regression
LD	Linkage disequilibrium
MAE	Mean absolute error
MSE	Mean squared error
PCA	Principal component analysis
r^2	Squared correlation coefficient
RF	Random forest
SNP	Single-nucleotide polymorphism
SVR	Support vector regression
WGS	Whole-genome sequencing/resequencing
XGBoost	Extreme gradient boosting
λ	Genomic inflation factor

References

1. Liang, X.F.; Oku, H.; Ogata, H.Y.; Liu, J.; He, X. Weaning Chinese Perch *Siniperca chuatsi* (Basilewsky) onto Artificial Diets based upon its specific sensory modality in feeding: Weaning Chinese Perch onto Artificial Diets. *Aquacult. Res.* **2001**, *32*, 76–82. [CrossRef]
2. He, S.; Li, L.; Lv, L.Y.; Cai, W.J.; Dou, Y.Q.; Li, J.; Tang, S.L.; Chen, X.; Zhang, Z.; Xu, J.; et al. Mandarin fish (Sinipercidae) genomes provide insights into innate predatory feeding. *Commun. Biol.* **2020**, *3*, 361. [CrossRef] [PubMed]
3. Sun, C.F.; Dong, J.J.; Li, W.H.; Tian, Y.Y.; Hu, J.; Ye, X. Response to four generations of selection for growth performance traits in mandarin fish (*Siniperca chuatsi*). *Aquaculture* **2022**, *548*, 737590. [CrossRef]
4. Bureau of Fisheries; Ministry of Agriculture and Rural Affairs; National Fisheries Technology; China Society of Fisheries. *China Fishery Statistical Yearbook*; China Agriculture Press: Beijing, China, 2026; pp. 24–25. ISBN 978-7-109-34659-8.
5. Xu, X.D.; Zhang, Y.P.; Ding, L.Y.; Huang, J.F.; Zhou, Z.Y.; Chen, W.J. Farmed Chinese Perch (*Siniperca chuatsi*) coinfected with parasites and oomycete pathogens. *Fishes* **2024**, *9*, 97. [CrossRef]
6. Guan, W.Z.; Qiu, G.F.; Qiu, G.F. Transcriptome analysis of the growth performance of hybrid mandarin fish after food conversion. *PLoS ONE* **2020**, *15*, e0240308. [CrossRef] [PubMed]
7. Ding, L.Y.; Zhang, Y.P.; Chen, J.C.; Chen, W.J.; Xie, S.Q.; Chen, Q.T. Growth, muscle nutrition composition, and digestive enzyme activities of the juvenile and adult *Siniperca chuatsi* fed on live baits and a formulated diet. *Fishes* **2022**, *7*, 379. [CrossRef]
8. Li, L.; He, S.; Lin, M.H.; Zhang, Y.P.; Kuhl, H.; Liang, X.F. Whole-genome resequencing and bisulfite sequencing provide new insights into the feeding habit domestication in mandarin fish (*Siniperca chuatsi*). *Front. Genet.* **2023**, *13*, 1088081. [CrossRef] [PubMed]
9. Geng, C.F.; Song, H.L.; Zhang, X.; Li, S.; Xia, Y.L.; Dong, Y.; Liu, S.; Su, Y.Q.; Wang, W.; Bai, S.; et al. Integrated genomic and transcriptomic analysis reveals candidate genes for *Siniperca chuatsi* adaptation to formulated diet. *Aquacult. Rep.* **2026**, *48*, 103658. [CrossRef]
10. He, S.; You, J.J.; Liang, X.F.; Zhang, Z.L.; Zhang, Y.P. Transcriptome sequencing and metabolome analysis of food habits domestication from live prey fish to artificial diets in mandarin fish (*Siniperca chuatsi*). *BMC Genom.* **2021**, *22*, 129. [CrossRef] [PubMed]
11. Luo, J.X.; Gao, X.T.; Rong, Z.; Zhang, L.H.; Sun, Y.F.; Qi, Z.L.; Yu, Q.; Waiho, K.; Zhao, W.X.; Xu, Y.H.; et al. Transcriptome sequencing reveals effects of artificial feed domestication on intestinal performance and gene expression of carnivorous mandarin fish (*Siniperca chuatsi*) and related mechanisms. *Mar. Biotechnol.* **2025**, *27*, 41–53. [CrossRef] [PubMed]
12. Chen, X.; Sun, C.F.; Dong, J.J.; Li, W.H.; Tian, Y.Y.; Hu, J.; Ye, X. Comparative analysis of the gut microbiota of mandarin fish (*Siniperca chuatsi*) feeding on compound diets and live baits. *Front. Genet.* **2022**, *13*, 797420. [CrossRef] [PubMed]
13. Yuan, Y.; Huang, Y.Z.; Li, H.; Li, R.B.; Wang, F.B.; Wang, Y.S. Comparative analysis of brain transcriptomics, intestinal metabolomics and intestinal microbial diversity between two body weight-differentiated groups of mandarin fish (*Siniperca chuatsi*) after artificial feed acclimation. *BMC Genom.* **2026**, *27*, 206. [CrossRef] [PubMed]
14. Yuan, Y.Y.; Peng, D.Z.; Gu, X.R.; Gong, Y.Z.; Sheng, Z.Y.; Hu, X.X. Polygenic basis and variable genetic architectures contribute to the complex nature of body weight —A genome-wide study in four Chinese indigenous chicken breeds. *Front. Genet.* **2018**, *9*, 229. [CrossRef] [PubMed]
15. Wright, K.M.; Deighan, A.G.; Di Francesco, A.; Freund, A.; Jojic, V.; Churchill, G.A.; Raj, A. Age and diet shape the genetic architecture of body weight in diversity outbred mice. *eLife* **2022**, *11*, e64329. [CrossRef] [PubMed]
16. Tu, G.X.; Zhang, X.S.; Jiang, R.R.; Zhang, L.; Lai, C.J.; Yan, Z.Y.; Lv, Y.R.; Weng, S.P.; Zhang, L.; He, J.G.; et al. Long-read genome assemblies reveal a *cis*-regulatory landscape associated with phenotypic divergence in two sister *Siniperca* fish species. *Zool. Res.* **2023**, *44*, 287–302. [CrossRef] [PubMed]
17. Ding, W.D.; Zhang, X.H.; Zhao, X.M.; Jing, W.; Cao, Z.M.; Li, J.; Huang, Y.; You, X.X.; Wang, M.; Shi, Q.; et al. A chromosome-level genome assembly of the mandarin fish (*Siniperca chuatsi*). *Front. Genet.* **2021**, *12*, 671650. [CrossRef] [PubMed]
18. Zhou, Y.; Fu, H.C.; Wang, Y.Y.; Huang, H.Z. Genome-wide association study reveals growth-related SNPs and candidate genes in mandarin fish (*Siniperca chuatsi*). *Aquaculture* **2022**, *550*, 737879. [CrossRef]
19. Liu, D.R.; Zhang, J.; Zou, Z.J.; Long, C.; Lin, J.Q.; Zeng, J.Y.; Hou, J.P.; Han, L.Q.; Jiang, Y.L.; Li, S.S.; et al. Identification of SNPs and candidate genes associate with growth performance in all-female mandarin fish (*Siniperca chuatsi*) by a genome-wide association study. *Aquaculture* **2024**, *586*, 740778. [CrossRef]
20. Ji, Z.H.; Zhang, Q.W.; Liao, C.L.; Huang, X.X.; Yu, H.D.; Xiong, Y.J.; Li, Y.X.; Xue, J.J.; Chen, F.F.; Zhang, X.Z. A comparative analysis of serum biochemistry, gene expression, and gut microbiota in domestication-susceptible and domestication-resistant mandarin fish (*Siniperca chuatsi*). *Fish Physiol. Biochem.* **2026**, *52*, 89. [CrossRef] [PubMed]
21. Sun, C.F.; Yan, Y.Y.; Zhang, Y.; Dong, J.J.; Chen, X.; Gao, F.Y.; Zhang, H.T.; Wang, Z.Y.; Wu, C.B.; Ye, X.; et al. Genome loci with selection signatures revealed in a domesticated mandarin fish population. *BMC Genom.* **2025**, *26*, 947. [CrossRef] [PubMed]
22. Yang, Y.H.; Liang, X.F.; Fang, R.; Peng, M.Y.; Huang, Z.D. Polymorphism of SNPs in the lipoprotein lipase (*LPL*) in *Siniperca chuatsi* and their association with feed habit domestication. *Hereditas* **2011**, *33*, 996–1002. [CrossRef] [PubMed]

23. Manthei, K.A.; Patra, D.; Wilson, C.J.; Fawaz, M.V.; Piersimoni, L.; Shenkar, J.C.; Yuan, W.M.; Andrews, P.C.; Engen, J.R.; Schwendeman, A.; et al. Structural analysis of lecithin: Cholesterol acyltransferase bound to high density lipoprotein particles. *Commun. Biol.* **2020**, *3*, 28. [CrossRef] [PubMed]
24. Nyame, K.; Xiong, J.; Alsohybe, H.N.; de Jong, A.P.H.; Peña, I.V.; de Miguel, R.; Brummelkamp, T.R.; Hartmann, G.; Nijman, S.M.B.; Raaben, M.; et al. PLA2G15 Is a BMP hydrolase and its targeting ameliorates lysosomal disease. *Nature* **2025**, *642*, 474–483. [CrossRef] [PubMed]
25. Matsuda, K.; Maruyama, K.; Miura, T.; Uchiyama, M.; Shioda, S. Anorexigenic action of pituitary adenylate cyclase-activating polypeptide (PACAP) in the goldfish: Feeding-induced changes in the expression of mRNAs for PACAP and its receptors in the brain, and locomotor response to central injection. *Neurosci. Lett.* **2005**, *386*, 9–13. [CrossRef] [PubMed]
26. Matsuda, K.; Kojima, K.; Shimakura, S.I.; Wada, K.; Maruyama, K.; Uchiyama, M.; Kikuyama, S.; Shioda, S. Corticotropin-releasing hormone mediates a-melanocyte-stimulating hormone-induced anorexigenic action in goldfish. *Peptides* **2008**, *29*, 1930–1936. [CrossRef] [PubMed]
27. Nakamachi, T.; Tanigawa, A.; Konno, N.; Shioda, S.; Matsuda, K. Expression patterns of PACAP and PAC1R genes and anorexigenic action of PACAP1 and PACAP2 in zebrafish. *Front. Endocrinol.* **2019**, *10*, 227. [CrossRef] [PubMed]
28. Zhang, Z.; Yuan, X.P.; Wu, H.; Gao, J.W.; Wu, J.Y.; Xiong, Z.Z.; Feng, Z.F.; Xie, M.; Li, S.M.; Xie, Z.G.; et al. The effect of short-Term artificial feed domestication on the expression of oxidative-stress-related genes and antioxidant capacity in the liver and gill tissues of mandarin fish (*Siniperca chuatsi*). *Genes* **2024**, *15*, 487. [CrossRef] [PubMed]
29. Escartín Pérez, R.E.; Mancilla Díaz, J.M.; Cortés Salazar, F.; López Alonso, V.E.; Florán Garduño, B. Chapter 5—CB1/5-HT/GABA Interactions and Food Intake Regulation. In *5-HT Interaction with Other Neurotransmitters: Experimental Evidence and Therapeutic Relevance—Part A*; Di Giovanni, G., De Deurwaerdere, P., Eds.; Progress in Brain Research; Elsevier: Amsterdam, The Netherlands, 2021; Volume 259, pp. 177–196. [CrossRef] [PubMed]
30. Han, Y.; Xia, G.B.; He, Y.L.; He, Y.; Farias, M.; Xu, Y.; Wu, Q. A hindbrain dopaminergic neural circuit prevents weight gain by reinforcing food satiation. *Sci. Adv.* **2021**, *7*, eabf8719. [CrossRef] [PubMed]
31. Yan, X.; Qin, C.B.; Yang, G.K.; Deng, D.P.; Yang, L.P.; Feng, J.C.; Mi, J.L.; Nie, G.X. The regulatory role of Apelin on the appetite and growth of common carp (*Cyprinus Carpio* L.). *Animals* **2020**, *10*, 2163. [CrossRef] [PubMed]
32. Volkoff, H.; Wyatt, J.L. Apelin in Goldfish (*Carassius auratus*): Cloning, distribution and role in appetite regulation. *Peptides* **2009**, *30*, 1434–1440. [CrossRef] [PubMed]
33. Valenti, M.; Cottone, E.; Martinez, R.; De Pedro, N.; Rubio, M.; Viveros, M.P.; Franzoni, M.F.; Delgado, M.J.; Di Marzo, V. The endocannabinoid system in the brain of *Carassius auratus* and its possible role in the control of food intake. *J. Neurochem.* **2005**, *95*, 662–672. [CrossRef] [PubMed]
34. Chivite, M.; Comesaña, S.; Calo, J.; Soengas, J.L.; Conde-Sieira, M. Endocannabinoid receptors are involved in enhancing food intake in rainbow trout. *Horm. Behav.* **2022**, *146*, 105277. [CrossRef] [PubMed]
35. Zhang, Z.B.; Wang, A.; Wang, Q.Y.; Gao, S.Q.; Wang, L.L.; Hu, H.H.; Nanaei, H.A.; Shah, A.M.; Liu, G.L.; Zhu, K.; et al. Cost-effective genomic prediction for fertility traits: A comparison of machine learning and conventional models using low-coverage sequencing in Holstein heifers. *Animal* **2026**, *20*, 101827. [CrossRef] [PubMed]
36. Song, H.L.; Wang, W.; Dong, T.; Yan, X.Y.; Geng, C.F.; Bai, S.; Hu, H.X. Prioritized SNP selection from whole-genome sequencing improves genomic prediction accuracy in sturgeons using linear and machine learning models. *Int. J. Mol. Sci.* **2025**, *26*, 7007. [CrossRef] [PubMed]
37. Tsairidou, S.; Hamilton, A.; Robledo, D.; Bron, J.E.; Houston, R.D. Optimizing low-cost genotyping and imputation strategies for genomic selection in Atlantic salmon. *G3 Genes Genomes Genet.* **2020**, *10*, 581–590. [CrossRef] [PubMed]
38. Fraslin, C.; Robledo, D.; Kause, A.; Houston, R.D. Potential of low-density genotype imputation for cost-efficient genomic selection for resistance to *Flavobacterium columnare* in rainbow trout (*Oncorhynchus mykiss*). *Genet. Sel. Evol.* **2023**, *55*, 59. [CrossRef] [PubMed]
39. Speiser, J.L.; Wolf, B.J.; Chung, D.; Karvellas, C.J.; Koch, D.G.; Durkalski, V.L. BiMM forest: A random forest method for modeling clustered and longitudinal binary outcomes. *Chemom. Intell. Lab. Syst.* **2019**, *185*, 122–134. [CrossRef] [PubMed]
40. Yoshida, G.M.; Bangera, R.; Carvalheiro, R.; Correa, K.; Figueroa, R.; Lhorente, J.P.; Yáñez, J.M. Genomic prediction accuracy for resistance against *Piscirickettsia salmonis* in farmed rainbow trout. *G3 Genes Genomes Genet.* **2018**, *8*, 719–726. [CrossRef] [PubMed]
41. McCarty, A.J.; Allen, S.K.; Plough, L.V. Genome-wide analysis of acute low salinity tolerance in the eastern oyster *Crassostrea virginica* and potential of genomic selection for trait improvement. *G3 Genes Genomes Genet.* **2022**, *12*, jkab368. [CrossRef] [PubMed]
42. Omeka, W.K.M.; Liyanage, D.S.; Lee, S.; Udayantha, H.M.V.; Kim, G.; Ganeshalingam, S.; Jeong, T.; Jones, D.B.; Massault, C.; Jerry, D.R.; et al. Genomic prediction model optimization for growth traits of olive flounder (*Paralichthys olivaceus*). *Aquacult. Rep.* **2024**, *36*, 102132. [CrossRef]

43. Thanh Vu, N.T.; Phuc, T.H.; Oanh, K.T.P.; Sang, N.V.; Trang, T.T.; Nguyen, N.H. Accuracies of genomic predictions for disease resistance of striped catfish to *Edwardsiella ictaluri* using artificial intelligence algorithms. *G3 Genes Genomes Genet.* **2022**, *12*, jkab361. [CrossRef] [PubMed]
44. Luo, Z.; Yu, Y.; Bao, Z.M.; Li, F.H. Evaluation of machine learning method in genomic selection for growth traits of Pacific white shrimp. *Aquaculture* **2024**, *581*, 740376. [CrossRef]
45. Xu, H.; Jin, Y.N.; Huang, R.Y.; Lv, Y.G.; Jin, J.; Wang, M.Y.; Ye, X.J.; Lai, M.Y.; Zhang, D.L. Estimation of genetic parameters for growth traits and genomic selection study in *Siniperca scherzeri*. *Mar. Biotechnol.* **2026**, *28*, 8. [CrossRef] [PubMed]
46. Wang, J.B.; Zhang, Z.W. GAPIT Version 4: Integration of GWAS into Genomic Prediction. *Mol. Biol. Evol.* **2026**, *43*, msag107. [CrossRef] [PubMed]
47. Huang, J.J.; Han, L.; Liang, J.; Liu, H.; Wu, Y.; Xu, P.; Xie, Z. An Efficient Method for Acclimating Mandarin Fish to Artificial Compound Feed. CN202211586169.0, 21 March 2023.
48. Niu, J.; Vasemägi, A.; López, M.; Pukk, L.; Huss, M.; Gårdmark, A. DNA extracted from boiled archival fish bones yields high quality whole genome sequencing data. *J. Fish Biol.* **2026**, *108*, 2137–2147. [CrossRef] [PubMed]
49. Chen, S.F.; Zhou, Y.Q.; Chen, Y.R.; Gu, J. fastp: An ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* **2018**, *34*, i884–i890. [CrossRef] [PubMed]
50. Li, H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv* **2013**, arXiv:1303.3997. [CrossRef]
51. Li, H.; Handsaker, B.; Wysoker, A.; Fennell, T.; Ruan, J.; Homer, N.; Marth, G.; Abecasis, G.; Durbin, R. 1000 genome project data processing subgroup. The sequence alignment/map format and SAMtools. *Bioinformatics* **2009**, *25*, 2078–2079. [CrossRef] [PubMed]
52. Tarasov, A.; Vilella, A.J.; Cuppen, E.; Nijman, I.J.; Prins, P. Sambamba: Fast processing of NGS alignment formats. *Bioinformatics* **2015**, *31*, 2032–2034. [CrossRef] [PubMed]
53. Browning, S.R.; Browning, B.L. Rapid and accurate haplotype phasing and missing-data inference for whole-genome association studies by use of localized haplotype clustering. *Am. J. Hum. Genet.* **2007**, *81*, 1084–1097. [CrossRef] [PubMed]
54. Chang, C.C.; Chow, C.C.; Tellier, L.C.; Vattikuti, S.; Purcell, S.M.; Lee, J.J. Second-generation PLINK: Rising to the challenge of larger and richer datasets. *Gigascience* **2015**, *4*, 7. [CrossRef] [PubMed]
55. Yang, J.; Lee, S.H.; Goddard, M.E.; Visscher, P.M. GCTA: A tool for genome-wide complex trait analysis. *Am. J. Hum. Genet.* **2011**, *88*, 76–82. [CrossRef] [PubMed]
56. Devlin, B.; Roeder, K. Genomic control for association studies. *Biometrics* **1999**, *55*, 997–1004. [CrossRef] [PubMed]
57. Zhang, C.; Dong, S.S.; Xu, J.Y.; He, W.M.; Yang, T.L. PopLDdecay: A fast and effective tool for linkage disequilibrium decay analysis based on variant call format files. *Bioinformatics* **2019**, *35*, 1786–1788. [CrossRef] [PubMed]
58. Zhou, X.; Stephens, M. Genome-wide efficient mixed-model analysis for association studies. *Nat. Genet.* **2012**, *44*, 821–824. [CrossRef] [PubMed]
59. Yin, L.L.; Zhang, H.H.; Tang, Z.S.; Xu, J.Y.; Yin, D.; Zhang, Z.W.; Yuan, X.H.; Zhu, M.J.; Zhao, S.H.; Li, X.Y.; et al. rMVP: A Memory-efficient, visualization-enhanced, and parallel-accelerated tool for genome-wide association study. *Genom. Proteom. Bioinform.* **2021**, *19*, 619–628. [CrossRef] [PubMed]
60. Cantalapiedra, C.P.; Hernández-Plaza, A.; Letunic, I.; Bork, P.; Huerta-Cepas, J. eggNOG-mapper v2: Functional annotation, Orthology assignments, and domain prediction at the metagenomic scale. *Mol. Biol. Evol.* **2021**, *38*, 5825–5829. [CrossRef] [PubMed]
61. Yu, G.C.; Wang, L.G.; Han, Y.Y.; He, Q.Y. clusterProfiler: An R package for comparing biological themes among gene clusters. *OMICS J. Integr. Biol.* **2012**, *16*, 284–287. [CrossRef] [PubMed]
62. Benjamini, Y.; Hochberg, Y. Controlling the false discovery rate: A practical and powerful approach to multiple testing. *J. R. Stat. Soc. B* **1995**, *57*, 289–300. [CrossRef]
63. VanRaden, P.M. Efficient methods to compute genomic predictions. *J. Dairy Sci.* **2008**, *91*, 4414–4423. [CrossRef] [PubMed]
64. Meuwissen, T.H.E.; Hayes, B.J.; Goddard, M.E. Prediction of total genetic value using genome-wide dense marker maps. *Genetics* **2001**, *157*, 1819–1829. [CrossRef] [PubMed]
65. Habier, D.; Fernando, R.L.; Kizilkaya, K.; Garrick, D.J. Extension of the Bayesian Alphabet for genomic selection. *BMC Bioinf.* **2011**, *12*, 186. [CrossRef] [PubMed]
66. De Los Campos, G.; Naya, H.; Gianola, D.; Crossa, J.; Legarra, A.; Manfredi, E.; Weigel, K.; Cotes, J.M. Predicting quantitative traits with regression models for dense molecular markers and pedigree. *Genetics* **2009**, *182*, 375–385. [CrossRef] [PubMed]
67. Drucker, H.; Burges, C.J.C.; Kaufman, L.; Smola, A.; Vapnik, V. Support vector regression machines. In *Proceedings of the 9th International Conference on Neural Information Processing Systems (NIPS 1996), Denver, CO, USA, 2–5 December 1996*; Mozer, M.C., Jordan, M.I., Petsche, T., Eds.; MIT Press: Cambridge, MA, USA, 1997; pp. 155–161.
68. Breiman, L. Random Forests. *Mach. Learn.* **2001**, *45*, 5–32. [CrossRef]

69. Saunders, C.; Gammerman, A.; Vovk, V. Ridge regression learning algorithm in dual variables. In *Proceedings of the 15th International Conference on Machine Learning (ICML 1998), Madison, WI, USA, 24–27 July 1998*; Morgan Kaufmann: San Francisco, CA, USA, 1998; pp. 515–521.
70. Chen, T.; Guestrin, C. XGBoost: A scalable tree Boosting System. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*; ACM: San Francisco, CA, USA, 2016; pp. 785–794. [CrossRef]
71. Madsen, P.; Jensen, J.; Labouriau, R.; Christensen, O.F.; Sahana, G. DMU—A Package for Analyzing Multivariate Mixed Models in Quantitative Genetics and Genomics. In Proceedings of the 10th World Congress of Genetics Applied to Livestock Production, Vancouver, BC, Canada, 17–22 August 2014; pp. 22–24.
72. Pérez, P.; De Los Campos, G. Genome-wide regression and prediction with the BGLR statistical package. *Genetics* **2014**, *198*, 483–495. [CrossRef] [PubMed]
73. Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B. Scikit-learn: Machine learning in python. *J. Mach. Learn. Res.* **2011**, *12*, 2825–2830. [CrossRef]
74. Long, N.; Gianola, D.; Rosa, G.J.M.; Weigel, K.A. Application of support vector regression to genome-assisted prediction of quantitative traits. *Theor. Appl. Genet.* **2011**, *123*, 1065–1074. [CrossRef] [PubMed]
75. Douak, F.; Melgani, F.; Benoudjit, N. Kernel ridge regression with active learning for wind speed prediction. *Appl. Energy* **2013**, *103*, 328–340. [CrossRef]
76. Schrauf, M.F.; De Los Campos, G.; Munilla, S. Comparing genomic prediction models by means of cross validation. *Front. Plant Sci.* **2021**, *12*, 734512. [CrossRef] [PubMed]
77. Gianola, D.; Schön, C.C. Cross-validation without doing cross-validation in genome-enabled prediction. *G3 Genes Genomes Genet.* **2016**, *6*, 3107–3128. [CrossRef] [PubMed]
78. Waldmann, P. On the use of the Pearson correlation coefficient for model evaluation in genome-wide prediction. *Front. Genet.* **2019**, *10*, 899. [CrossRef] [PubMed]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.