Genomic Prediction for Whole Weight, Body Shape, Meat Yield, and Color Traits in the Portuguese Oyster *Crassostrea angulata*

Sang V. Vu[1,2,3]*, Wayne Knibb[1], Cedric Gondro[4]*, Sankar Subramanian[1,2], Ngoc T. H. Nguyen[3], Mobashwer Alam[5], Michael Dove[6], Arthur R. Gilmour[7], In Van Vu[3], Salma Bhyan[1,2], Rick Tearle[8], Le Duy Khuong[9], Tuan Son Le[10] and Wayne O'Connor[1,2,6]*

[1] GeneCology Research Centre, University of the Sunshine Coast, Sippy Downs, QLD, Australia, [2] School of Science, Technology and Engineering, University of the Sunshine Coast, Sippy Downs, QLD, Australia, [3] Northern National Broodstock Center for Mariculture, Research Institute for Aquaculture Number 1, Hai Phong, Vietnam, [4] Department of Animal Science, College of Agriculture and Natural Resources, Michigan State University, East Lansing, MI, United States, [5] Queensland Alliance for Agriculture and Food Innovation, The University of Queensland, Saint Lucia, QLD, Australia, [6] NSW Department of Primary Industries, Port Stephens Fisheries Institute, Taylors Beach, NSW, Australia, [7] Consultant, Orange, NSW, Australia, [8] School of Animal and Veterinary Science, The University of Adelaide, Adelaide, SA, Australia, [9] Faculty of Environment, Ha Long University, Uong Bi, Vietnam, [10] Research Institute for Marine Fisheries, Ngo Quyen, Hai Phong, Vietnam

Genetic improvement for quality traits, especially color and meat yield, has been limited in aquaculture because the assessment of these traits requires that the animals be slaughtered first. Genotyping technologies do, however, provide an opportunity to improve the selection efficiency for these traits. The main purpose of this study is to assess the potential for using genomic information to improve meat yield (soft tissue weight and condition index), body shape (cup and fan ratios), color (shell and mantle), and whole weight traits at harvest in the Portuguese oyster, *Crassostrea angulata*. The study consisted of 647 oysters: 188 oysters from 57 full-sib families from the first generation and 459 oysters from 33 full-sib families from the second generation. The number per family ranged from two to eight oysters for the first and 12–15 oysters for the second generation. After quality control, a set of 13,048 markers were analyzed to estimate the genetic parameters (heritability and genetic correlation) and predictive accuracy of the genomic selection for these traits. The multi-locus mixed model analysis indicated high estimates of heritability for meat yield traits: 0.43 for soft tissue weight and 0.77 for condition index. The estimated genomic heritabilities were 0.45 for whole weight, 0.24 for cup ratio, and 0.33 for fan ratio and ranged from 0.14 to 0.54 for color traits. The genetic correlations among whole weight, meat yield, and body shape traits were favorably positive, suggesting that the selection for whole weight would have beneficial effects on meat yield and body shape traits. Of paramount importance is the fact that the genomic prediction showed moderate to high accuracy for the traits studied (0.38–0.92). Therefore, there are good prospects to improve whole weight, meat yield, body shape, and color traits using genomic information. A multi-trait selection program using the genomic information can boost the genetic gain and minimize inbreeding in the long-term for this population.

Keywords: genetic improvement, genomic selection, high density SNP markers, prediction accuracy, genetic parameters

INTRODUCTION

The Portuguese oyster, *Crassostrea angulata*, is becoming an important aquaculture mollusk species in various parts of the world, including Asia and Europe (Grade et al., 2016; Vu et al., 2017a; Gagnaire et al., 2018). The hatchery and production technologies are well established for this species. A genetic improvement program was established in Vietnam in 2014 using stock from three hatcheries to form the base population (Vu et al., 2017b). Initially, the breeding program focused solely on increasing the whole weight at harvest using traditional family selection approaches. The genetic evaluation of the first three generations (generation interval is 1 year) of the program showed a considerable genetic gain of around 6.0% per generation in the whole weight at harvest (Vu et al., 2019). Other traits such as meat yield, body shape ("cup" and "fan" ratios), and color of the shell and mantle have more recently been evaluated and incorporated into the breeding program due to an increasing market demand for these traits (Vu et al., 2019, 2020). Seafood color, especially for oysters that are usually sold live or in the half-shell, affects consumers' preferences and their willingness to buy because certain colors are perceived to be associated with better flavor and quality (Kahn and Wansink, 2004; Xing et al., 2018). In addition, shell shape has been reported to be related to meat yield in mollusk species (Gimin et al., 2004). Misshapen animals fetch a lower price than well-formed oysters (Lee et al., 2004). Previous studies showed that the heritability for meat yield, body shape, and color traits varied from 0.13 to 0.57 and has potential for genetic improvement (Vu et al., 2019, 2020). Conventional selection, however, is less efficient for traits that cannot be measured in living animals (Meuwissen et al., 2013). This is the case for meat yield and mantle color as these traits require slaughter before measurements can be taken. The current breeding program relies on measurements of meat yield and mantle color from the siblings of the selection candidates and on correlated traits. As a result, selection accuracy has been low and genetic progress has been slow for these traits. An alternative would be to add genomic selection for these difficult-to-measure traits to the breeding program; it is considered a more efficient way to speed up the rate of genetic progress and increase the prediction accuracy (Muranty et al., 2014).

Genomic selection is now widely used in many agricultural species, and it can significantly improve traits that are determined by many loci with small effects (Robledo et al., 2018b). Genomic selection can be used to estimate the breeding values of individuals with no pedigree information and can more accurately predict breeding values than traditional selection based on pedigree records because it uses genetic markers to build a genomic relationship matrix to be used to estimate the breeding values for individuals that better approximate *realized* relationships (Goddard and Hayes, 2007). The advantages of genomic selection over traditional pedigree-based approaches in terms of the accuracy of the predictions for polygenic traits have been well established for aquaculture species (Vallejo et al., 2017; Gutierrez et al., 2018, 2019; Hollenbeck and Johnston, 2018; Joshi et al., 2019). In addition, genomic selection is extremely useful for exploring within-family variability to increase the rates of genetic

gain and to reduce generation intervals (Georges et al., 2019). In aquaculture species, the accuracy of genomic prediction reported for growth, carcass, and meat quality traits as well as disease resistance traits ranged between 0.16 and 0.83 (Castillo-Juárez et al., 2015; Zenger et al., 2018; Joshi et al., 2019; Liu et al., 2019; Yoshida et al., 2019). Genomic prediction has not yet been widely adopted for mollusk breeding, but there are a few examples of its use for traits such as growth and disease resistance in Pacific oysters, *Crassostrea gigas* (Gutierrez et al., 2018, 2019), and pearl quality in pearl oysters, *Pinctada maxima* (Jones et al., 2017). To our knowledge, there are no reports on the genomic prediction for whole weight in Portuguese oysters nor for meat yield (soft tissue weight and condition index), body shape, and color (shell and mantle) in any other oyster species.

The purpose of this study was to (i) estimate the genomic heritability and genetic correlations for meat yield, color traits, harvest whole weight, and body shape and (ii) evaluate the accuracy of genomic prediction for these traits in Portuguese oysters.

MATERIALS AND METHODS

Oyster Samples

The Portuguese oysters used in this study originated from a breeding program to improve the whole weight at harvest in Vietnam (Vu et al., 2019). The oysters were produced in a hatchery and then raised in the open ocean from spat (2–4 mm) until harvest. The oyster families were separately raised in the ocean. Tissue samples were collected from the oysters at harvest after about a 9-month period of culture. Information on the animals included sire, dam, spawning date, harvest date, and rearing condition. Phenotypic measurements taken on the oysters included whole weight, body shape, meat yield, and color traits. Tissue samples were collected from a total of 647 oysters: 188 oysters representing 57 full-sib families from the first generation and 459 oysters representing 33 full-sib families from the second generation. The number per family ranged from two to eight oysters for the first and 12–15 oysters for the second generation. Most of the oysters from the second generation were progeny of the first generation of oysters (Vu et al., 2021b). All tissue samples were preserved in 80% ethanol, stored at −80°C, and sent to Diversity Array Technology Pty. Ltd., Canberra, Australia, for sequencing.

Traits Studied
Growth
The oyster shells were cleaned in water before taking the measurements. The whole weight of oysters at harvest was recorded using electronic scales with an accuracy of 0.01 g.

Body Shape
The cup and fan ratios were used as indicators of the body shape of the oysters. The cup ratio was calculated by dividing the shell width by the shell depth, and the fan ratio was calculated by dividing the shell length by the shell depth (Walton et al., 2013).

Meat Yield

Electronic scales with an accuracy of 0.01 g were used to measure the soft tissue weight. The condition index was calculated as soft tissue weight multiplied by 100/(whole weight minus shell weight) (Lawrence and Scott, 1982).

Color

The shell and mantle colors of oysters at harvest were measured using a FRU colorimeter WR10 8 mm model (Giang et al., 2019) *via* a CIELab L*a*b colorimeter which determines the color as a number under a device-independent 3D color model (Leon et al., 2006). The three dimensions in the model are L* (whiteness), a* (redness), and b* (yellowness). The values of L* range from 0 for black to 100 for white; a* values are negative for green and positive for red; and b* values are negative for blue and positive for yellow (b* = zero, neutral color) (Giang et al., 2019).

DNA Extraction, Library Construction, and Genotyping

The frozen oyster mantle samples were sequenced using DArTSeq^{TM} (Diversity Arrays Technology Pty. Ltd., Canberra, Australia) at a high marker density of 2.5 million sequences per sample. Genomic DNA was extracted and purified by Diversity Array Technology Pty. Ltd. (**Supplementary Table S1**). The SNP development and analysis were as described by Kilian et al. (2012). Briefly, DNA samples were digested with *Pst*I–*Sph*I (Kilian et al., 2012; Vu et al., 2021a) and then ligated using two different adaptors with two different restriction enzyme overhangs. The *Pst*I-compatible adapter consists of Illumina's flow-cell attachment, a sequencing primer, and a "staggered" varying-length barcode region (Elshire et al., 2011). The reverse adapter included a flow-cell attachment region and a *Sph*I-compatible overhang sequence. The PCR reaction used to amplify the mixed fragments of *Pst*I–*Sph*I following the reaction conditions comprised of an initial denaturation for 1 min at 94°C, followed by 30 cycles of 94°C for 20 s, 58°C for 30 s, and 72°C for 45 s, and a final extension at 72°C for 7 min.

The PCR products from each sample were fed to the c-Bot (Illumina) bridge PCR, followed by sequencing 77 nucleotide single-end reads on an Illumina Hiseq 2500. Reads with inaccurate barcode sequences were filtered out. Approximately 2,500,000 sequences per barcode/sample were used in marker calling. Finally, identical sequences were collapsed into "fastqcoll files," which were "groomed" to correct for low-quality bases. The "groomed" fastqcoll files were used in a secondary pipeline incorporating DArT SNP and SilicoDArT (presence/absence of restriction fragments in representation), called analysis algorithm DArTsoft14, which clustered all tags from all libraries at a threshold of 3 for SNP calling. Technical parameters, especially the balance of read counts for the allelic pairs, were used to parse into separate SNP loci. Additional selection criteria included analysis of approximately 1,000 controlled cross-populations and testing for Mendelian distribution of alleles to assign consistency scores to classify high-quality/low-error rate markers. Calling quality was assured by a high average read depth per locus (the average across all markers was over 30 reads/locus).

The sample and marker statistics are given in **Supplementary Tables S2,S3**, respectively. There were 18,849 SNP markers identified from genotyping by sequencing. Quality control was conducted with a call rate of 50% of SNP presence in the samples, leaving 13,048 SNP markers retained for analysis. The missing genotypes were imputed using the SVS Suite software with a default setting, changing them to average values (Bozeman, 2010).

Estimation of Narrow Sense Heritability and Genetic Correlations

Univariate linear mixed models were used to estimate the narrow sense heritabilities of the traits. The fixed effects fitted in the model were generation (age is a linear covariate nested within generation), sex determined at harvest (age based on spawning and harvest dates), and their interactions. The additive genetic effect and the residuals were used as random factors.

Heritability was calculated as the ratio of σ_a^2 on σ_p^2, where σ_a^2 is the additive genetic variance and σ_p^2 is the phenotypic variance calculated as $\sigma_p^2 = \sigma_a^2 + \sigma_e^2$ and σ_e^2 is the residual variance.

Genetic correlations between traits were estimated using a bivariate version of the same model (Mrode, 2014). All analyses were conducted using the average information restricted maximum likelihood procedures in the SVS Suite software (Bozeman, 2010).

Genomic Prediction

The accuracy of genomic prediction was estimated for all traits by fivefold cross-validation analysis (training set—80% and validation set—20%) with five replicates for each trait using gBLUP through the SVS Suite software (Bozeman, 2010). The gBLUP method assumes equal distribution and variance for individual SNP effects. The genome-wide complex trait analysis method was used to calculate the normalized genomic relationship matrix in our analysis (Yang et al., 2011). The gBLUP model is written in matrix notation as follows:

$$y = m + X\beta + Ma + e$$

where y is the vector of phenotypic observations (whole weight, body shape, meat yield, and color traits), m = overall mean, and X is the incidence matrix consisting of fixed effects (generation, sex, age, and interactions) in β. The matrix M is the incidence matrix of genetic effects, and the genetic values are $g = Ma$, such that $g_{n \times 1} \sim N(0n, G\sigma_g^2)$ and the genomic relationship matrix is $G_{n \times n} = \frac{MM'}{\sum_{j=1}^{m} 2p_j q_j}$ (VanRaden, 2008), in which $p_j's$ are the minor allele frequencies of the SNP genotypes ($q_j = 1 - p_j$) and $\sigma_g^2 = \sum_{j=1}^{m} 2p_j q_j \sigma_a^2$. Using the genomic best linear unbiased prediction (GBLUP) (VanRaden, 2008), genetic values were then fitted (on discovery population) and predicted (on validation population) by solving the mixed model equation.

The allele substitution effects (ASE) and fixed effect coefficients obtained from iterations and k-folds of the cross-validation analysis that gave the largest R^2-value were used to

predict the phenotypes of individual animals with the following model:

$$\hat{y} = X\hat{\beta} + M\hat{\alpha}$$

where $\hat{y}$ is the predicted phenotypes, X is the fixed effects matrix, $\hat{\beta}$ is the fixed effect coefficient, M is the genotype matrix, and $\hat{\alpha}$ is the ASE values.

The fivefold cross-validation using the gBLUP method was conducted using the SVS Suite (Bozeman, 2010). The actual and predicted phenotypes were compared using linear regression analysis. The coefficient of determination (R^2) from the regression analysis was used to assess the predictive ability of the gBLUP model for the traits studied. Furthermore, the animal effects were also predicted by traditional BLUP using the pedigree-based numerator relationship matrix for comparison with the genomic-based predictions described above. The pedigree summary of the oyster population in this study is given in **Supplementary Table S4**.

Population Genetics

Nucleotide diversity was estimated using the method developed by Nei and Li (1979). Using an in-house computer program, the genotype file obtained from DArTSeq was converted to the variable call format (VCF). This file was used as an input for the online software *VCF2PopTree* (Subramanian et al., 2019) to obtain the pairwise divergence matrix of all combinations of oyster individuals. This matrix was then used to construct a neighbor-joining tree using the software MEGA (Kumar et al., 2018). Fixation index (F_{ST}) was calculated using the nucleotide diversities of the total and sub-populations (Nei, 1973).

RESULTS

Phenotypic Data

The summary statistics of the traits are shown in **Table 1**. The average whole weight at harvest calculated from the 647 oysters in this study was 51.07 g, with the data range within three standard deviations of the mean (**Table 1**). There was considerable variation in whole weight, with a wide range from 15.71 to 69.90 g. The shell and mantle color had a significant yellow component (mean b* significantly greater than zero) and was dark (L around 25% but with considerable variation). The average cup and fan ratios indicated a typical depth/width/length ratio of 1:1.6:2.9. The largest variation levels were also found in color traits of tissues and shells, especially for shell color a*, with values from −2.07 to 54.96, and mantle color L*, from 7.24 to 63.70.

Heritability and Genetic Correlation

The genomic heritability (h^2), the proportion of total variance explained by the SNP markers, ranged from 0.14 to 0.77 (**Table 2**). The heritability of the shell color traits ranged from moderate to high (0.14–0.54), and the estimates of heritability were 0.13 for a* (red), 0.26 for b* (yellow), and 0.34 for L* (lightness). For the mantle, these traits had heritabilities of 0.54 for red, 0.49 for yellow, and 0.34 for lightness. The heritability estimates for meat yield traits were relatively high (0.43–0.77). Similarly, the estimate of heritability for whole weight was quite high at 0.45. However, the heritability estimates for body shape using the indicators of cup and fan ratios were low to moderate at 0.24 and 0.33, respectively.

Genetic Correlations

The positive genetic correlation between whole weight and soft tissue weight was high at 0.63, suggesting that selection for improved whole weight can result in desirable changes in soft tissue weight. However, low or insignificant genetic correlations were observed between whole weight and the other traits: color traits (−0.14 to 0.34), condition index (0.18), and body shape (0.08–0.14), implying that the selection for growth would result in little or no changes in these traits in this population. In addition, the genetic correlations between whole weight and all color traits were not only small, but they all had large standard errors and were therefore not significant. The genomic genetic correlations (r_g) of whole weight with the other traits are presented in the **Table 2**.

Accuracy of Genomic Prediction

Table 3 shows the prediction accuracy of genomic selection for the traits using the fivefold cross-validation method. High prediction accuracies using GBLUP were found for meat yield and color traits, where measurements cannot be taken on living animals. The accuracies were 0.71 for soft tissue weight and 0.49 for condition index. For the color traits, the accuracies ranged between 0.38 and 0.92. The GBLUP prediction accuracy was comparatively high for the harvest whole weight at 0.74; similarly, the body shape traits also showed a high accuracy, with 0.40 for cup ratio and 0.63 for fan ratio. Across all traits, the GBLUP predictions had higher prediction accuracies than the traditional BLUP prediction accuracies.

The association test of predicted and actual phenotypes for four predominant traits are presented in **Figure 1A** for the whole weight, **Figure 1B** for soft tissue, **Figure 1C** for mantle color a*, and **Figure 1D** for mantle color b*. The R^2 values varied from 0.37 for the mantle color L* to 0.91 for the shell color a*.

DISCUSSION

Our results have made five important discoveries that can be used to improve the genetic progress of the current breeding program for Portuguese oysters. Specifically, they answer the following questions.

Is There Any Genetic Variation in the Traits Studied?

Our analysis indicated a large genetic variation, with estimates of genomic heritability falling in the range of 0.14–0.77. Our results were somewhat higher than those we have previously obtained with the traditional pedigree-based relationship matrix in the same population (Vu et al., 2019). The genomic heritability for meat yield and body shape, in particular, was much larger than that reported using a pedigree-based estimation (Vu et al., 2019).

TABLE 1 | Summary of the phenotypic data.

Groups	Traits	Unit	n	Mean	SD	Min	Max
Growth	Whole weight	g	647	51.07	16.13	15.71	69.90
Body shape	Cup ratio	Ratio	361	1.61	0.33	0.74	2.60
	Fan ratio	Ratio	361	2.93	0.72	1.46	5.07
Meat yield	Soft tissue weight	g	490	9.55	3.13	1.05	19.92
	Condition index	Index	284	67.78	21.65	14.29	99.93
Color	Shell color a*	CIELab	306	2.35	3.71	−2.07	54.96
	Shell color b*	CIELab	306	11.58	4.25	3.99	43.02
	Shell color L*	CIELab	306	23.63	12.30	6.49	69.71
	Mantle color a*	CIELab	448	4.40	3.46	0.16	24.60
	Mantle color b*	CIELab	448	10.11	4.01	1.02	23.30
	Mantle color L*	CIELab	448	24.21	8.73	7.24	63.70

TABLE 2 | SNP heritability (h^2) and genetic correlations (r_g) of whole weight with other traits.

Groups	Traits	Heritability		Genetic correlation	
		h^2	SE	r_g	SE
Growth	Whole weight	0.45	0.06		
Body shape	Cup ratio	0.24	0.10	0.14	0.24
	Fan ratio	0.33	0.08	0.08	0.15
Meat yield	Soft tissue weight	0.43	0.07	0.63	0.10
	Condition index	0.77	0.07	0.18	0.15
Color	Shell color a*	0.14	0.07	−0.13	0.22
	Shell color b*	0.26	0.09	−0.02	0.13
	Shell color L*	0.34	0.09	0.34	0.19
	Mantle color a*	0.54	0.06	−0.13	0.13
	Mantle color b*	0.49	0.07	−0.14	0.15
	Mantle color L*	0.34	0.09	0.14	0.19

TABLE 3 | Genomic prediction accuracy of the traits with GBLUP and traditional BLUP.

Factors		Single variate linear mixed model	
Groups	Traits	GBLUP	BLUP
Growth	Whole weight	0.74	0.66
Body shape	Cup ratio	0.40	0.29
	Fan ratio	0.63	0.55
Meat yield	Soft tissue weight	0.71	0.51
	Condition index	0.49	0.22
Color	Shell color a*	0.92	0.46
	Shell color b*	0.62	0.65
	Shell color L*	0.83	0.66
	Mantle color a*	0.87	0.67
	Mantle color b*	0.89	0.44
	Mantle color L*	0.38	0.14

Similarly, the estimates of heritability based on SNP markers for mantle color traits (h^2 = 0.34–0.54) were around twice those based on pedigree (h^2 = 0.15–0.33) (Vu et al., 2020). These differences between genomic and pedigree heritability estimates of meat yield, body shape, and color traits are most likely due to errors in pedigree recording between spawning and harvesting. These results suggest that genomic selection can be extremely efficient for traits that cannot be measured directly on the selection candidates especially since oyster breeding involves a few large families, and pedigree information is difficult to record accurately. Unfortunately, there have been no other studies using a genomic selection approach to estimate heritability for meat yield, color, and body shape traits in mollusk species that could be used for comparison. Finally, the estimation of heritability for whole weight in this study was significantly higher than that reported in the Pacific oyster by Gutierrez et al. (2018) (h^2 = 0.45
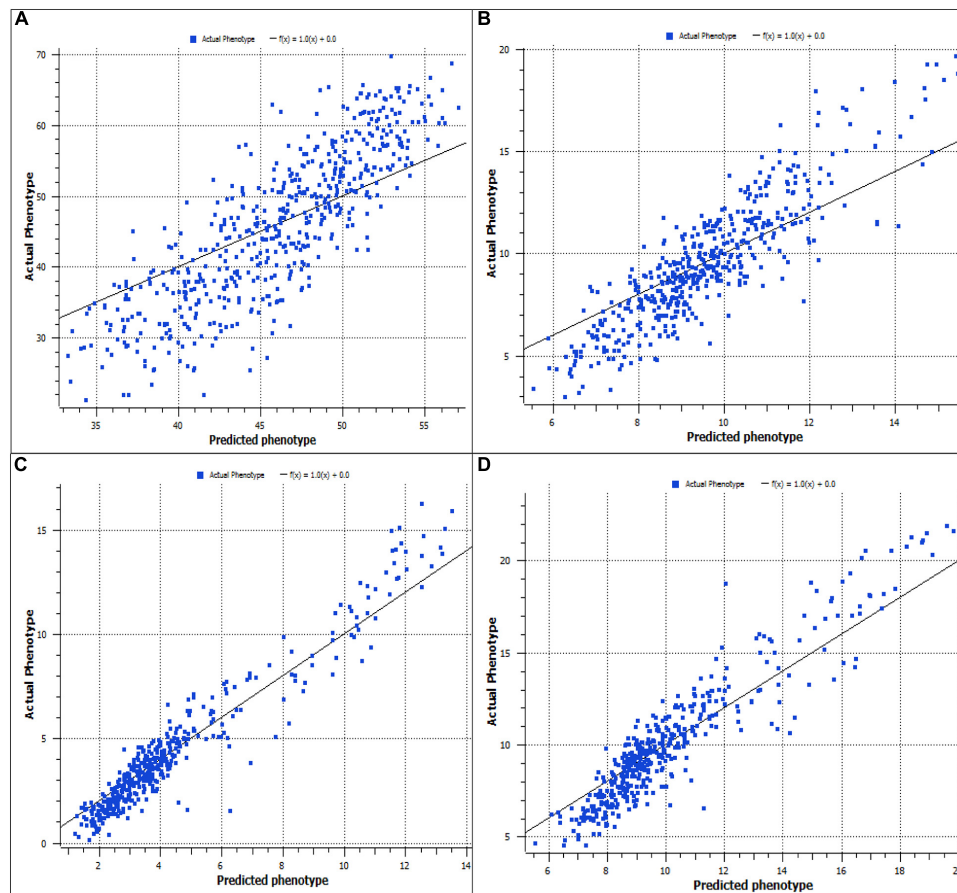
FIGURE 1 | Prediction accuracy for whole weight **(A)**, soft tissue weight **(B)**, mantle color a **(C)**, and mantle color b **(D)**.

vs. 0.35, respectively). However, estimated heritability in genomic selection can also depend on species, sample size, marker density, relationship of reference and validation population, and statistical methods for analysis (Georges et al., 2019). It should be noted that our population is highly inter-related, and the high estimates probably only pertain to selection within that population.

Does This Genomic Analysis Predict That Selection for Harvest Whole Weight Will Affect Body Shape, Meat Yield, and Color?

The genetic correlations obtained with the GBLUP analysis were generally similar to those obtained from the traditional BLUP method (Vu et al., 2019). The high and positive genetic correlation between harvest whole and soft tissue weight indicates that these two traits are controlled by some common sets of genes. This result is in line with the previous estimate by Vu et al. (2019) using the BLUP method (0.63 for GBLUP vs. 0.50 for BLUP). Therefore, soft tissue weight will increase along with the harvest whole weight under the selection for harvest whole weight. Meanwhile, the small positive genetic

correlations found between harvest whole weight and body shape show that there is some/limited potential to improve these traits by selection for harvest whole weight, and any genetic progress will be very slow. Therefore, trait groups such as whole weight, body shape, and meat yield can improve in the same direction, and this suggests that the selection for one trait (harvest whole weight) will lead to an improvement in these other traits. In contrast, the negative and positive but low genetic correlations between whole weight and color traits found in this study agree with those reported in the same population using the pedigree BLUP method (Vu et al., 2020). Consequently, our results suggest that no potential exists to improve color traits by selection for harvest whole weight. Taken together, this study provides fundamental information to better understand the genetic architecture of quantitative complex traits in Portuguese oysters.

Is Genomic Selection Reliable for the Traits Studied?

Our study reports, for the first time, the predictive accuracy of genomic selection for color and meat yield traits in a mollusk species. The predictive accuracies using the GBLUP model for meat yield and color traits were significantly higher than those
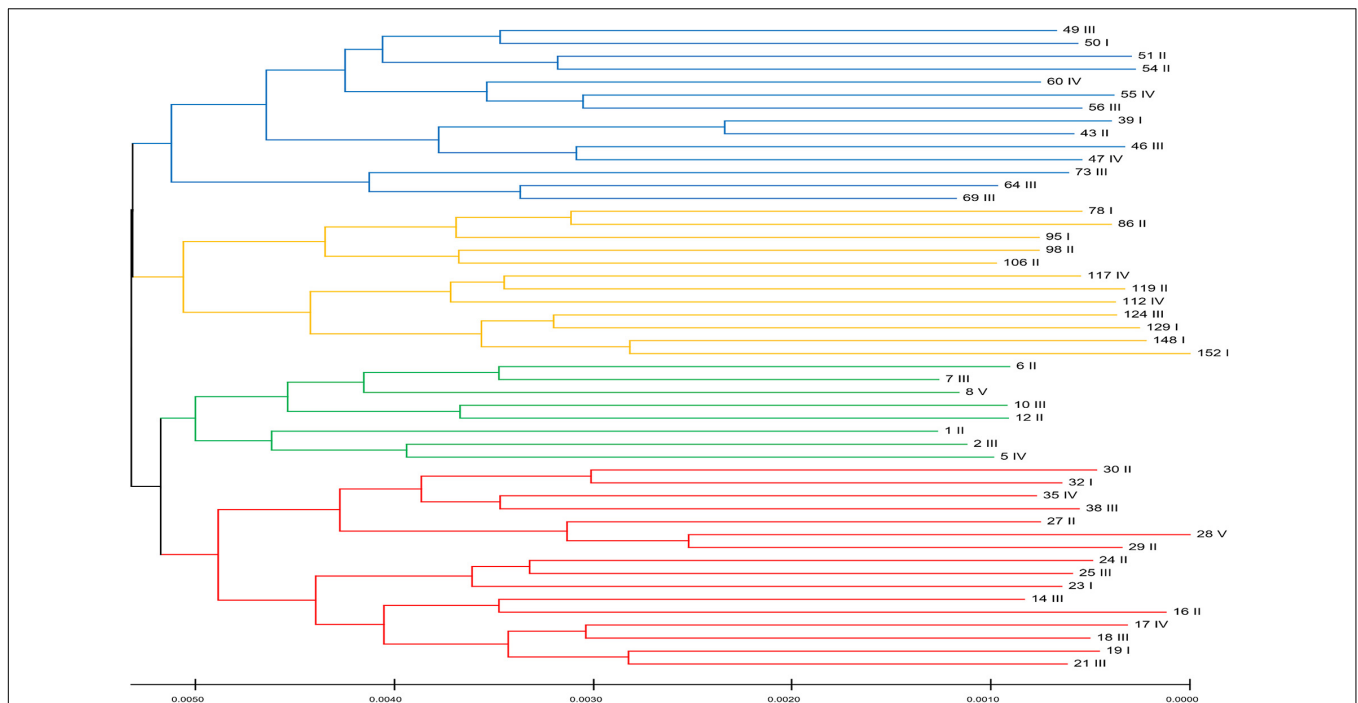
FIGURE 2 | The structure of the population for this Portuguese oyster population in the first generation. Colors represent four distinct clusters based on genetic similarities. Roman numerals represent the whole weight category that the individual belong to: I: 30–40, II: 40–50, III: 50–60, IV: 60–70, and V: >70 g. This figure demonstrates that the oysters in the population do not cluster based on their whole-body weight but only based on their genetic relationship.

obtained from the pedigree BLUP model, suggesting that the genomic selection for these traits is more efficient than the pedigree-based selection. The predictive accuracies using the GBLUP method for meat yield and color traits are in good agreement with those reported in banana shrimp (Nguyen et al., 2019). However, these accuracies were higher than those obtained for growth-related traits in the Pacific oyster (Gutierrez et al., 2018), Pacific white leg shrimp (Wang et al., 2017), and yellow drum (Liu et al., 2019). In addition, the accuracies from our study are high due to the close relationships of the sampled oysters. The differences among the studies could originate from the SNP marker density, the set of SNP markers used for the analysis (Liu et al., 2019), the training population size (Andonov et al., 2017; Zenger et al., 2018; Georges et al., 2019), the relationship between the training and validation datasets (Hayes and Goddard, 2001), the relationships between individuals in the reference population (Hayes et al., 2009), and the heritability of the traits (Daetwyler et al., 2008; Georges et al., 2019). To the best of our knowledge, no studies have reported the prediction accuracy of genomic selection for body shape in any other aquaculture species. Collectively, across aquaculture species, our estimates of accuracies for whole weight, body shape, meat yield, and color traits in Portuguese oysters fall in the range observed in other aquaculture species (Zenger et al., 2018). The high levels of prediction accuracy open new selection opportunities to improve meat yield and color traits in the Portuguese oyster.

Does Population Stratification Affect the Results of This Study?

To quantify the genetic variation in the data, we estimated the nucleotide diversity, which produced a value of 0.009, suggesting 0.9% variation in the population. This estimate was slightly less than that observed for the Pacific oyster (1.2%) (Zhang et al., 2012). One of the confounding factors in genomic selection method is the structure of the populations, owing to the effects of genetic drift (Goddard and Hayes, 2007). Some of the genetic markers can, by chance alone, correlate with the traits due to the familial or pedigree structure of the populations compared. To examine the genetic relatedness among the first-generation animals, we selected 50 individuals, excluding their siblings. We categorized these oysters based on their whole-body weight as I: 30–40, II: 40–50, III: 50–60, IV: 60–70, and V: >70 g. We then constructed a neighbor-joining tree (**Figure 2**), which revealed four distinct clusters (four colors) in the population based on their genetic similarities. The fixation index (F_{ST}) of 0.13 among these clusters suggested the existence of a population structure. However, it is evident from **Figure 2**, which shows that the whole weight of the members of each cluster varies significantly within a cluster. Therefore, these results suggest that the population structure observed in the data does not correlate with the phenotypic trait (whole weight) evaluated in this study. Hence, population stratification is unlikely to influence the results of this study.

Are There Any Prospects for Using Genomic Selection in This Portuguese Oyster Population?

Genomic selection programs can be improved by using updated estimates of genetic parameters and benefiting from the higher predictive accuracy for quantitative complex traits, especially meat yield and color traits, in the Portuguese oyster. The favorably positive genetic correlations between whole weight, body shape, and meat yield traits found in this study suggest that there is little need to use a selection index to improve these related traits. To further illustrate, the heritability and genetic correlations from our study were used in SelAction (Rutten et al., 2002) to simulate the selection process and indicated that the traits whole weight, meat yield, and body shape can be improved simultaneously. The selection for enhancing whole weight results in beneficial changes in the other traits. However, it is necessary to use a selection index to improve color traits as well as weight traits due to the low genetic correlations between whole weight and color traits.

Collectively, the results from this study showed that there are prospects for the application of genomic selection to improve a range of complex traits in the Portuguese oyster. Consequently, further studies should be carried out to collect further information on predictive power and potential genetic gain. Firstly, the training population size needs to be increased to obtain a more accurate prediction (Goddard, 2009), i.e., we need to increase our sample size from the rather small population of 647 oysters used in this study to a much larger reference population which represents more families and generations. One strategy can be a mixture of low- and high-density panels, where the broodstock is sequenced at a higher marker density and the offspring are sequenced at a lower density, and then the lower density panels are imputed up to the higher density for the genomic prediction (Yoshida et al., 2018; Tsairidou et al., 2020). This strategy allows a larger number of samples to be sequenced, thus a larger training population size. This has been done at medium density in Pacific oysters for growth-related traits or in salmon (Gutierrez et al., 2018). Secondly, the use of a genome-wide association study—informative SNPs instead of random SNPs—was able to assist in improving the predictive accuracy (Liu et al., 2019). Therefore, the use of a smaller number of informative SNPs may help to reduce the cost of genotyping in this population. Thirdly, the prediction equations were evaluated in the reference population in which both phenotypes and genotypes were available; validation should be carried out in a breeding population where the selection candidates are the training population, and validation is on subsequent generations. Fourthly, individual grouping methods such as mean weight of family based on genomic information should be considered in Portuguese oysters to improve the accuracy for further studies. These individual grouping methods have been shown to increase the predictive accuracy and resulted in a higher genetic gain (Chu et al., 2019).

From the discussion above, the optimization of genomic selection will reduce the generation interval and increase the genetic gain in this population. Lowering the generation interval may result in an increase in inbreeding rate in the long-term.

Therefore, to balance between inbreeding rate and genetic gain, genetic improvement programs have applied optimal genomic selection methodology and strategies (Nielsen et al., 2011; Gorjanc and Hickey, 2018; Robledo et al., 2018a). For the Portuguese oyster sector, if genomic selection is integrated into the breeding program, it could reduce the generation interval below 1 year per generation. The lower generation interval contributes to a faster life cycle, resulting in a higher economic revenue due to saving labor and time. In addition, estimates of breeding values using genomic information are more accurate than those using physical tags in the pedigree, where errors may occur during spawning, larval rearing, or pedigree management. Therefore, genomic selection will allow a more accurate selection of oysters to become parents in the next generations. However, a major impediment to widely using genomic selection in the aquaculture industry is the cost of genotyping and phenotyping the selection candidates. The former can be dealt with by using low-density SNP panels that can balance between prediction accuracy and sequencing costs to achieve the accuracy needed for genomic selection (Lillehammer et al., 2013; Gutierrez et al., 2018; Tsairidou et al., 2020; Boudry et al., 2021; Vu et al., 2021a).

CONCLUSION

The predictive accuracy using genomic selection was relatively high for all the traits studied. In addition, the estimates of heritability in the meat yield traits such as soft tissue weight and condition index were high, while those of color traits such as shell and mantle color were from low to high. The genomic selection for improvement of whole weight leads to desirable changes in other traits, such as meat yield and body shape traits, but will not affect color traits. Future breeding programs should combine all traits into a selection index to bring about higher revenues for aquaculture farmers and entrepreneurs.

DATA AVAILABILITY STATEMENT

The DNA sequence data used in this study is available and can be accessed through the USC Research Bank under doi 10.25907/00038. The phenotypic data is available by request from the corresponding authors.

ETHICS STATEMENT

There are no requirements on ethics approval for mollusk species in Vietnam.

AUTHOR CONTRIBUTIONS

SV, WK, and WO'C conceived and designed the experiments and selected the samples for sequencing. SV and NN prepared, drafted, and edited the manuscript. SV, CG, MA, and SB analyzed the data. SS involved in the population structure analysis. TL and LK prepared the oyster tissue samples. SS, CG, MA, MD, AG, RT, IV, and WO'C participated in editing the manuscript. All authors read and approved the manuscript.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2021.661276/full#supplementary-material

REFERENCES

Andonov, S., Lourenco, D., Fragomeni, B., Masuda, Y., Pocrnic, I., Tsuruta, S., et al. (2017). Accuracy of breeding values in small genotyped populations using different sources of external information—a simulation study. *J. Dairy Sci.* 100, 395–401. doi: 10.3168/jds.2016-11335

Boudry, P., Allal, F., Aslam, M. L., Bargelloni, L., Bean, T. P., Brard-Fudulea, S., et al. (2021). Current status and potential of genomic selection to improve selective breeding in the main aquaculture species of international council for the exploration of the sea (ICES) member countries. *Aquac. Rep.* 20:100700. doi: 10.1016/j.aqrep.2021.100700

Bozeman, M. (2010). *Golden Helix, Inc. SNP & Variation Suite (Version 7. x).*

Castillo-Juárez, H., Campos-Montes, G. R., Caballero-Zamora, A., and Montaldo, H. H. (2015). Genetic improvement of Pacific white shrimp *[Penaeus (Litopenaeus) vannamei]*: perspectives for genomic selection. *Front. Genet.* 6:93.

Chu, T. T., Bastiaansen, J. W., Berg, P., and Komen, H. (2019). Optimized grouping to increase accuracy of prediction of breeding values based on group records in genomic selection breeding programs. *Genet. Sel. Evol.* 51:64.

Daetwyler, H. D., Villanueva, B., and Woolliams, J. A. (2008). Accuracy of predicting the genetic risk of disease using a genome-wide approach. *PLoS One* 3:e3395. doi: 10.1371/journal.pone.0003395

Elshire, R. J., Glaubitz, J. C., Sun, Q., Poland, J. A., Kawamoto, K., Buckler, E. S., et al. (2011). A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PLoS One* 6:e19379. doi: 10.1371/journal.pone.0019379

Gagnaire, P.-A., Lamy, J.-B., Cornette, F., Heurtebise, S., Dégremont, L., Flahauw, E., et al. (2018). Analysis of genome-wide differentiation between native and introduced populations of the cupped oysters *Crassostrea gigas* and *Crassostrea angulata. Genome Biol. Evol.* 10, 2518–2534. doi: 10.1093/gbe/evy194

Georges, M., Charlier, C., and Hayes, B. (2019). Harnessing genomic information for livestock improvement. *Nat. Rev. Genet.* 20, 135–156. doi: 10.1038/s41576-018-0082-2

Giang, C. T., Knibb, W., Ninh, N. H., and Nguyen, N. H. (2019). Prospects for genetic improvement in objective measurements of body colour in Pacific Whiteleg Shrimp (*Litopenaeus vannamei*). *J. Mar. Sci. Eng.* 7:460. doi: 10.3390/jmse7120460

Gimin, R., Mohan, R., Thinh, L., and Griffiths, A. (2004). The relationship of shell dimensions and shell volume to live weight and soft tissue weight in the mangrove clam, *Polymesoda erosa* (Solander, 1786) from northern Australia. *NAGA WorldFish Center Quarterly* 27, 32–35.

Goddard, M. (2009). Genomic selection: prediction of accuracy and maximisation of long term response. *Genetica* 136, 245–257. doi: 10.1007/s10709-008-9308-0

Goddard, M., and Hayes, B. (2007). Genomic selection. *J. Anim. Breed. Genet.* 124, 323–330.

Gorjanc, G., and Hickey, J. M. (2018). AlphaMate: a program for optimizing selection, maintenance of diversity and mate allocation in breeding programs. *Bioinformatics* 34, 3408–3411. doi: 10.1093/bioinformatics/bty375

Grade, A., Chairi, H., Lallias, D., Power, D. M., Ruano, F., Leitao, A., et al. (2016). New insights about the introduction of the Portuguese oyster, *Crassostrea angulata*, into the North East Atlantic from Asia based on a highly polymorphic mitochondrial region. *Aquat. Living Resour.* 29:404. doi: 10.1051/alr/2016035

Gutierrez, A. P., Matika, O., Bean, T. P., and Houston, R. D. (2018). Genomic selection for growth traits in Pacific oyster (*Crassostrea gigas*): potential of low-density marker panels for breeding value prediction. *Front. Genet.* 9:391.

Gutierrez, A. P., Symonds, J., King, N., Steiner, K., Bean, T. P., and Houston, R. D. (2019). Potential of genomic selection for improvement of resistance to Ostreid Herpes virus in Pacific oyster (*Crassostrea gigas*). *BioRxiv [Preprint]* doi: 10.1101/754473

Hayes, B. J., Visscher, P. M., and Goddard, M. E. (2009). Increased accuracy of artificial selection by using the realized relationship matrix. *Genet. Res.* 91, 47–60. doi: 10.1017/s0016672308009981

Hayes, B., and Goddard, M. (2001). Prediction of total genetic value using genome-wide dense marker maps. *Genet. Sel. Evol.* 157, 1819–1829. doi: 10.1093/genetics/157.4.1819

Hollenbeck, C. M., and Johnston, I. A. (2018). Genomic tools and selective breeding in molluscs. *Front. Genet.* 9:253.

Jones, D., Toole, P., Khatkar, M., Raadsma, H., Jerry, D., and Zenger, K. (2017). Developing a genomic selection breeding program for complex pearl colour traits within the silver-lipped pearl oyster. *Proc. Assoc. Advancement Anim. Breed. Genet.* 22, 537–540.

Joshi, R., Skaarud, A., de Vera, M., Alvarez, A. T., and Ødegård, J. (2019). Genomic prediction for commercial traits using univariate and multivariate approaches in Nile tilapia (*Oreochromis niloticus*). *Aquaculture* 516:734641. doi: 10.1016/j.aquaculture.2019.734641

Kahn, B. E., and Wansink, B. (2004). The influence of assortment structure on perceived variety and consumption quantities. *J. Consum. Res.* 30, 519–533. doi: 10.1086/380286

Kilian, A., Wenzl, P., Huttner, E., Carling, J., Xia, L., Blois, H., et al. (2012). Diversity arrays technology: a generic genome profiling technology on open platforms. *Methods Mol. Biol.* 888, 67–89. doi: 10.1007/978-1-61779-870-2_5

Kumar, S., Stecher, G., Li, M., Knyaz, C., and Tamura, K. (2018). MEGA X: molecular evolutionary genetics analysis across computing platforms. *Mol. Biol. Evol.* 35, 1547–1549. doi: 10.1093/molbev/msy096

Lawrence, D., and Scott, G. (1982). The determination and use of condition index of oysters. *Estuaries* 5, 23–27. doi: 10.2307/1352213

Lee, D.-J., Xu, X., Lane, R. M., and Zhan, P. (2004). "Shape analysis for an automatic oyster grading system," in *Two-and Three-Dimensional Vision Systems for Inspection, Control, and Metrology II*, Vol. 5606, ed. K. G. Harding (Bellingham: International Society for Optics and Photonics).

Leon, K., Mery, D., Pedreschi, F., and Leon, J. (2006). Color measurement in L* a* b* units from RGB digital images. *Food Res. Int.* 39, 1084–1091. doi: 10.1016/j.foodres.2006.03.006

Lillehammer, M., Meuwissen, T. H., and Sonesson, A. K. (2013). A low-marker density implementation of genomic selection in aquaculture using within-family genomic breeding values. *Genet. Sel. Evol.* 45, 1–8.

Liu, G., Dong, L., Gu, L., Han, Z., Zhang, W., Fang, M., et al. (2019). Evaluation of genomic selection for seven economic traits in yellow drum (*Nibea albiflora*). *Mar. Biotechnol.* 21, 806–812. doi: 10.1007/s10126-019-09925-7

Meuwissen, T., Hayes, B., and Goddard, M. (2013). Accelerating improvement of livestock with genomic selection. *Annu. Rev. Anim. Biosci.* 1, 221–237. doi: 10.1146/annurev-animal-031412-103705

Mrode, R. A. (2014). *Linear Models for the Prediction of Animal Breeding Values.* Wallingfort: CABI.

Muranty, H., Jorge, V., Bastien, C., Lepoittevin, C., Bouffier, L., and Sanchez, L. (2014). Potential for marker-assisted selection for forest tree breeding: lessons from 20 years of MAS in crops. *Tree Genet.* 10, 1491–1510. doi: 10.1007/s11295-014-0790-5

Nei, M. (1973). Analysis of gene diversity in subdivided populations. *Proc. Natl. Acad. Sci. U. S. A.* 70, 3321–3323. doi: 10.1073/pnas.70.12.3321

Nei, M., and Li, W.-H. (1979). Mathematical model for studying genetic variation in terms of restriction endonucleases. *Proc. Natl. Acad. Sci. U. S. A.* 76, 5269–5273. doi: 10.1073/pnas.76.10.5269

Nguyen, N. H., Phuthaworn, C., and Knibb, W. (2019). Genomic prediction for disease resistance to Hepatopancreatic parvovirus and growth, carcass and quality traits in *Banana shrimp Fenneropenaeus merguiensis. Genomics* 112, 2021–2027. doi: 10.1016/j.ygeno.2019.11.014

Nielsen, H., Sonesson, A., and Meuwissen, T. (2011). Optimum contribution selection using traditional best linear unbiased prediction and genomic breeding values in aquaculture breeding schemes. *J. Anim. Sci.* 89, 630–638. doi: 10.2527/jas.2009-2731

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

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

Rutten, M., Bijma, P., Woolliams, J., and Van Arendonk, J. (2002). SelAction: software to predict selection response and rate of inbreeding in livestock breeding programs. *J. Heredity* 93, 456–458. doi: 10.1093/jhered/93.6.456

Subramanian, S., Ramasamy, U., and Chen, D. (2019). VCF2PopTree: a client-side software to construct population phylogeny from genome-wide SNPs. *PeerJ* 7:e8213.

Tsairidou, S., Hamilton, A., Robledo, D., Bron, J. E., and Houston, R. D. (2020). Optimizing low-cost genotyping and imputation strategies for genomic selection in Atlantic salmon. *G3* 10, 581–590. doi: 10.1534/g3.119.400800

Vallejo, R. L., Leeds, T. D., Gao, G., Parsons, J. E., Martin, K. E., Evenhuis, J. P., et al. (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.* 49:17.

VanRaden, P. M. (2008). Efficient methods to compute genomic predictions. *J. Dairy Sci.* 91, 4414–4423. doi: 10.3168/jds.2007-0980

Vu, S. V., Gondro, C., Nguyen, N. T. H., Gilmour, A. R., Tearle, R., Knibb, W., et al. (2021a). Prediction accuracies of genomic selection for nine commercially important traits in the Portuguese oyster (*Crassostrea angulata*) using DArT-Seq technology. *Genes* 12:210. doi: 10.3390/genes12020210

Vu, S. V., Knibb, W., O'Connor, W., Nguyen, N. T., Van In, V., Dove, M., et al. (2020). Genetic parameters for traits affecting consumer preferences for the Portuguese oyster, *Crassostrea angulata. Aquaculture* 526:735391. doi: 10.1016/j.aquaculture.2020.735391

Vu, S. V., Premachandra, H. K. A., O'Connor, W., Nguyen, N. T. H., Dove, M., Vu, I. V., et al. (2021b). Development of SNP parentage assignment in the Portuguese oyster *Crassostrea angulata. Aquac. Rep.* 19:100615. doi: 10.1016/j.aqrep.2021.100615

Vu, V. I., O'Connor, W., Vu, V. S., Phan, T. V., and Knibb, W. (2017a). Resolution of the controversial relationship between Pacific and Portuguese

oysters internationally and in Vietnam. *Aquaculture* 473, 389–399. doi: 10.1016/j.aquaculture.2017.03.004

Vu, V. I., Vu, V. S., O'Connor, W., Phan, T. V., Dove, M., Knibb, W., et al. (2017b). Are strain genetic effect and heterosis expression altered with culture system and rearing environment in the Portuguese oyster (*Crassostrea angulata*)? *Aquac. Res.* 48, 4058–4069. doi: 10.1111/are.13227

Vu, V. S., Knibb, W., Nguyen, T. H. N., Vu, V. I., O'Connor, W., Dove, M., et al. (2019). First breeding program of the Portuguese oyster *Crassostrea angulata* demonstrated significant selection response in traits of economic importance. *Aquaculture* 518:734664. doi: 10.1016/j.aquaculture.2019.734664

Walton, W., Rikard, F., Chaplin, G., Davis, J., Arias, C., and Supan, J. J. A. (2013). Effects of ploidy and gear on the performance of cultured oysters, *Crassostrea virginica*: survival, growth, shape, condition index and Vibrio abundances. *Aquaculture* 414, 260–266. doi: 10.1016/j.aquaculture.2013.07.032

Wang, Q., Yu, Y., Li, F., Zhang, X., and Xiang, J. (2017). Predictive ability of genomic selection models for breeding value estimation on growth traits of Pacific white shrimp *Litopenaeus vannamei. Chinese J. Oceanol. Limnol.* 35, 1221–1229. doi: 10.1007/s00343-017-6038-0

Xing, D., Li, Q., Kong, L., and Yu, H. (2018). Heritability estimate for mantle edge pigmentation and correlation with shell pigmentation in the white-shell strain of Pacific oyster, *Crassostrea gigas. Aquaculture* 482, 73–77. doi: 10.1016/j.aquaculture.2017.09.026

Yang, J., Lee, S. H., Goddard, M. E., and Visscher, P. M. (2011). GCTA: a tool for genome-wide complex trait analysis. *Am. J. Hum. Genet.* 88, 76–82. doi: 10.1016/j.ajhg.2010.11.011

Yoshida, G. M., Carvalheiro, R., Lhorente, J. P., Correa, K., Figueroa, R., Houston, R. D., et al. (2018). Accuracy of genotype imputation and genomic predictions in a two-generation farmed Atlantic salmon population using high-density and low-density SNP panels. *Aquaculture* 491, 147–154. doi: 10.1016/j.aquaculture.2018.03.004

Yoshida, G. M., Lhorente, J. P., Correa, K., Soto, J., Salas, D., and Yáñez, J. M. (2019). Genome-wide association study and cost-efficient genomic predictions for growth and fillet yield in Nile tilapia (*Oreochromis niloticus*). *G3* 9, 2597–2607. doi: 10.1534/g3.119.400116

Zenger, K. R., Khatkar, M. S., Jones, D. B., Khalilisamani, N., Jerry, D. R., and Raadsma, H. W. (2018). Genomic selection in aquaculture: application, limitations and opportunities with special reference to marine shrimp and pearl oysters. *Front. Genet.* 9:693.

Zhang, G., Fang, X., Guo, X., Li, L., Luo, R., Xu, F., et al. (2012). The oyster genome reveals stress adaptation and complexity of shell formation. *Nature* 490, 49–54.