Aquaculture International (2025) 33:551
https://doi.org/10.1007/s10499-025-02210-6

RESEARCH

Genome wide association study of shell growth, condition index, shell and mantle colour in the Portuguese oyster, *Crassostrea angulata*

Sang Van Vu[1] · Wayne O'Connor[2] · In Van Vu[3] · Cedric Gondro[4] · Thu Thi Anh Nguyen[5] · Shantanu Kundu[6,7] · Kim Hyun Woo[8,9] · Soo Rin Lee[8,10] · Tran Dang Khanh[11] · Tiep Khac Nguyen[12] · Hien Van Doan[13,14] · Hsu Htoo[8] · Almas A. Gheyas[15]

Received: 7 March 2025 / Accepted: 16 August 2025

Abstract
The Portuguese oyster (*Crassostrea angulata*) is a commercially significant aquaculture species, gaining rapid popularity, particularly in Asia. Growth-related traits, along with shell and mantle colouration, are key determinants of market value in farmed oysters. To support trait improvement in breeding programs, this study investigates the genetic architecture of economically important traits, including shell length, shell height, shell width, condition index, and shell and mantle colour using genome-wide association study (GWAS). Using DArTseq technology, 647 oyster samples from two generations of a breeding program in Vietnam were genotyped: 188 samples from 57 full-sib families in the first generation and 459 samples from 33 full-sib families in the second. GWAS identified 31 significant SNPs associated with various traits, 24 of which mapped to protein-coding genes. Notable candidate genes associated with growth traits included *CE128-like* (*LOC128177318*) and *WIPI3-like* (*LOC128167327*), implicated in protein localization and autophagy, respectively. For colour-related traits, key candidate genes included *glucose dehydrogenase* (*LOC128184820*), *Neurobeachin-like* (*LOC128156661*), and *POP1-like* (*LOC128164428*), which are linked to catalytic activities, membrane trafficking, and RNA processing, suggesting roles in pigmentation and biomineralization. Additionally, *Neo-calmodulin-like* (*LOC128183296*), a gene involved in calcium binding, was identified as a candidate for shell colour, consistent with findings in other oyster species. The small effect sizes of the significant SNPs across all traits suggest polygenic control, underscoring the potential of genomic selection for trait improvement. This study provides foundational insights to inform selective breeding programs aimed at enhancing growth and aesthetic traits in *C. angulata*, contributing to the sustainability and profitability of oyster aquaculture.

Keywords *Crassostrea angulata* · Portuguese oysters · GWAS · Growth traits · Colour traits

Handling Editor: Pierre Boudry

Extended author information available on the last page of the article

Introduction

Oyster farming provides a sustainable source of animal protein due to its low-investment production systems, reduced environmental footprint compared to other forms of animal farming, and numerous ecosystem benefits (NOAA Fisheries 2022; Brigida 2023). Rich in vitamins, minerals, and omega-3 fatty acids, oysters are highly nutritious and a popular seafood choice in many parts of the world (Woolmer n.d.). Beyond their nutritional value, oyster farming offers significant environmental and ecological benefits (Brigida 2023). As filter feeders, oysters remove excess organic matter and phytoplankton from the water column, improving water quality and helping to combat eutrophication. Oyster reefs create complex habitats that support diverse marine species, promoting biodiversity, and act as natural buffers against coastal erosion (Brigida 2023). Additionally, oysters contribute to carbon sequestration through shell formation, helping to mitigate the impact of excess carbon dioxide in aquatic environments (Ugalde et al. 2023). Oysters also exhibit high adaptability to farming conditions and require no supplemental feed, chemicals, or medicines, resulting in a significantly lower carbon footprint compared to the farming of many other animal species (Seafish 2020). These combined benefits are driving the rapid global expansion of oyster farming. The Portuguese oyster (*Crassostrea angulata*, also referred to as *Magallana angulata*), is a key aquaculture species in Asia, with China producing 2 million tons annually (Wu et al. 2023). Vietnam has also experienced significant growth in oyster farming over the past decade, with *C. angulata* playing an important role in this expansion (Ugalde et al. 2023). Its fast growth and high survival rate make it a preferred species for aquaculture, with current production in Vietnam estimated at 50,000 tons (Le et al. 2023).

Improving production efficiency is critical for the commercial success of aquaculture. Like other production systems, growth is also a crucial trait in oyster aquaculture, as faster growth reduces production costs by shortening cultivation cycles and thereby improves production efficiency. However, other traits, including condition index, mantle colour, and shell morphology, also hold significant economic potential. Mantle colour, for instance, is a key target for selective breeding as it influences consumer preferences and perceptions of flavour and quality (Kahn and Wansink 2004; Xing et al. 2018). In a recent study, shell and mantle colour traits exhibited heritability ranging from 0.14 to 0.54, indicating potential for genetic improvement through selection (Vu et al. 2021b). The condition index, which reflects meat yield relative to shell weight, impacts product quality and market value (Kause et al. 2011). Oysters with higher condition indices—indicating greater meat yield and lighter shell weight—are particularly desirable (O'Connor et al. 2019). High heritability (0.77) was observed for the condition index in a previous study on the Portuguese oyster (Vu et al. 2021b), suggesting that a strong selection response is possible. Similarly, shell morphology traits, such as length, width, and height, are directly linked to soft tissue yield, making them important determinants of oyster pricing.

Genetic and genomic approaches are increasingly being used in the aquaculture sector to enhance production efficiency and improve economically important traits. The growing availability of genomic data, enabled by affordable sequencing and genotyping technologies, is allowing many production systems to leverage genomics-driven improvements, such as marker-assisted selection and genomic selection, to enhance selection accuracy and boost efficiency (Houston et al. 2020). A key focus of genomics is elucidating genetic architecture of traits under selection and using this information for trait improvement. Various genomic approaches can be employed for this purpose, including quantitative trait loci (QTL) mapping, Genome-Wide Association Studies (GWAS), and transcriptomic analysis.

These methods have been widely applied in many aquaculture species to inform selective breeding programs (Houston et al. 2020; Ajithkumar et al. 2025).

GWAS using high-throughput genotype data have identified single nucleotide polymorphisms (SNPs) and genes linked to economically important traits in various mollusc species as well. For example, molecular markers have been associated with shell growth and shape in Pacific oysters, *Crassostrea gigas* (He et al. 2021), nacre colour and growth traits in Pearl oysters, *Pinctada fucata* (Wang et al. 2022; Zhao et al. 2023), cadmium accumulation in Fujian oysters, *C. angulata* (Wu et al. 2023), shell colour in Bay scallops, *Argopecten irradians irradians* (Zhu et al. 2021), and heat tolerance in Pacific abalone, *Haliotis discus hannai* (Yu et al. 2021). These studies indicate that the genetic architecture of such traits is often polygenic, involving numerous genes with small effects, or oligogenic, with a few moderate-to-large effect QTL. Exceptions include traits like growth in European oysters, *Ostrea edulis* (Peñaloza et al. 2022) and ostreid herpesvirus resistance in Pacific oysters, *C. gigas* (Divilov et al. 2019), which are influenced by fewer genomic regions or distinct QTLs. While high throughput genomic tools are widely used in other industries, their adoption in aquaculture remains limited, mainly due to the high cost of whole genome sequencing and SNP arrays, particularly for non-model species. Genotyping-by-sequencing (GBS) offers a cheaper alternative in such cases. Among these, DArTseq™ combines DArT-based complexity reduction with next-generation sequencing, providing customizable, cost-effective genotyping (as low as $20–35 USD per sample) with high marker density (Vu et al. 2021a).

Despite the economic importance of *C. angulata*, genomic studies in this species are still sparse. Recent studies have identified candidate genes and markers linked to some growth-related traits (Vu et al. 2024; Xie et al. 2024). However, no markers have been reported for other commercially significant traits such as shell growth, condition index, and shell or meat colour in *C. angulata*. This study aims to address this gap by identifying candidate genes and genetic variants associated with these traits, which, alongside other growth-related traits, are critical determinants of market value. Using genome-wide SNP data generated via DArTSeq genotype-by-sequencing, this study conducts a GWAS to explore trait-genotype associations in *C. angulata*.

Materials and methods

Sampled oysters

A total of 647 individual oyster samples were obtained from a *C. angulata* breeding program in Vietnam, as detailed in previous studies (Vu et al. 2020, 2021b, 2021c). These included 188 samples from 57 full-sib families in the first generation (G1) of the breeding program and 459 samples from 33 full-sib families in the second generation (G2). The first-generation families were produced from a base population generated through diallel crossing of broodstock from three different commercial hatcheries in Vietnam, representing a mixture of genetic backgrounds. The second generation was subsequently produced by interbreeding individuals from the first generation. Family size ranged from two to eight oysters in the first generation and from 12 to 15 oysters in the second generation. Oyster families were reared separately in tanks until they reached two months of age, when they were transferred to a common culture environment. Each oyster family was cultured on four strings, each containing 7–8 oyster shells, with approximately 20–25

oyster spat attached per shell. Each family was labeled on a separate string for identification and record-keeping. All oysters were reared under uniform environmental conditions within each generation. Phenotypic traits, including shell length, shell width, shell height, total weight, condition index, and shell and mantle coloration, were recorded at the harvest stage, approximately 11–12 months of age. Due to resource limitations, only growth-related traits were measured across both generations, whereas colour traits were only recorded in the second generation. Furthermore, variations in the number of samples available for GWAS among traits also arose from other factors, such as missing phenotypic data and the exclusion of samples subsequent to quality control (QC) on genotypic and phenotypic data. At the time of sampling the oysters reached their reproductive stage, ready for spawning. Males and females were of similar proportion in the collected samples (males = 313, of which 80 originated from G1 and 233 from G2; females = 334, of which 108 originated from G1 and 226 from G2). Mantle tissue samples were preserved in 80% ethanol and stored at − 80 °C until sequencing.

Traits and phenotypic measurements

Growth-related traits and condition index

Shell dimensions (length, width, and height) were measured using calipers with an accuracy of 0.01 cm, while shell weight was determined using a digital balance accurate to 0.01 g. The condition index was calculated using the following formula, adapted from the method described by Lawrence and Scott (1982) as follows:

$$\text{Condition Index} = (\text{dry soft tissue weight} \times 100)/(\text{whole weight} - \text{shell weight}) \quad (1)$$

Shell and mantle colour

The shell and tissue colours of oysters at harvest were measured using an FRU colourimeter WR10 8 mm model via a CIELab L*a*b colourimeter which determines the colour as a number under a device-independent 3D colour 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 colour) (Vu et al. 2021b).

DNA extraction, library construction and sequencing

DNA extraction, purification, library construction, sequencing and SNP calling were performed by Diversity Array Technology Pty Ltd. (DArT), Canberra, Australia, using a standardized protocol (Kilian et al. 2012). For a detailed description of the DNA extraction, library construction, and sequencing procedures for the same set of samples analysed here, see Vu et al. (2021b). Briefly, DArTSeq generated about 70 nucleotide single-end reads on an Illumina Hiseq 2500. Reads with inaccurate barcode sequences were removed from the fastq files using stringent filtration criteria for the barcode region (Kilian et al. 2012). Approximately 2,500,000 sequence reads per barcode/sample were generated and subsequently used for variant calling.

Sequence data analysis, SNP calling and quality control

The sequence data was mapped against the latest version of the chromosome-level genome assembly (*refSeq ASM2576567v3*) for *C. angulata* using two different aligners — BLASTn (Altschul et al. 1990) and BWA-MEM2 (Vasimuddin et al. 2019) to obtain a consensus result for SNP calling. This approach was taken to minimize the risk of mapping errors due to short read length of the sequences. The BLASTn alignment was performed with conservative parameters: gap open 20, gap extend 5, penalty −5, word size 40, percent identity 99, and ungapped alignment. Alignments with suboptimal scores, those mapping to multiple positions, or those not aligning to chromosomes were removed. Alignment with BWA-MEM2 was performed using default parameters, except for a minimum seed length (k) of 40 and re-seeding (r) of 2.5. The same filtering criteria were applied as above to remove sub-optimal alignments. Consensus results from both aligners yielded 10,137 mapped SNPs. Homozygous markers and those missing in more than 10% of samples were discarded, Leaving 10,116 markers. Sporadic missing markers were imputed using Eagle2 (Loh et al. 2016) with default parameters. A final minor allele frequency (MAF) filter of 1% was applied, resulting in a final SNP set of 9846 markers. A 1% MAF threshold, rather than the more commonly used 5%, was applied to retain greater number of genetic variations, as DArTSeq — a reduced representation approach — yields fewer SNPs than whole-genome sequencing. SNPs with MAF > 1% are generally considered common variants (Visscher et al. 2017), making this threshold appropriate for capturing informative markers while maintaining allele frequencies suitable for GWAS. Filtering based on HWE probability was not applied, as it may also exclude important SNPs significantly associated with traits (Fardo et al. 2009). SNPs were pruned for linkage disequilibrium (LD) to remove linked loci at the r^2 threshold > 0.20 using PriorityPruner version 0.1.4 (available at http://prioritypruner.sourceforge.net/). Individuals with missing genotypes in more than 50% of the SNPs were excluded from analysis.

Statistical methods for genome-wide association study (GWAS)

For the GWAS analysis, we used the genotypic data from 9846 genome-wide SNP markers and a genomic best linear unbiased prediction (GBLUP) method to estimate SNP effects by back solving the EBVs using GWASTools package (Gogarten et al. 2012). The mixed model used was of the form:

$$\mathrm{y} = \mathrm{Xb} + \mathrm{Zu} + \mathrm{e} \tag{2}$$

where $\boldsymbol{y}$ is the vector of phenotypic measurements (shell length, width, height and weight, condition index, and shell and mantle colour), $\boldsymbol{b}$ is the fixed effects vector (year-class, sex, and tank), $\boldsymbol{u}$ is the vector of random genetic effects, $\boldsymbol{e}$ is the vector of residual errors, and $\boldsymbol{X}$ and $\boldsymbol{Z}$ are the design matrices for the fixed and random effects. The distributional assumption of the random effects was multivariate normal with mean zero and $\mathrm{var}\begin{bmatrix}\alpha \\ e\end{bmatrix} = \begin{bmatrix} G\sigma_\alpha^2 & 0 \\ 0 & I\sigma_e^2 \end{bmatrix}$, where σ_α^2, and σ_e^2 are respectively the additive genetic and residual variances. $\boldsymbol{G}$ is the genomic relationship matrix obtained from the SNP markers, and $\boldsymbol{I}$ is an identity matrix. The $\boldsymbol{G}$ matrix was built as follows: $G = M'M/\sum_{j=1}^{m} 2p_jq_j$, where $p_j s$ are the minor allele frequencies of the SNP genotypes ($q_j = 1 - p_j$), and $\boldsymbol{M}$ the SNPs genotype matrix (VanRaden 2008). SNPs with p-value < 0.0001 were considered to have a significant effect.

Additionally, we computed the genomic inflation factor (lambda, λ) to assess the presence of inflation/deflation in the test statistics. Lambda was calculated by taking the median of the Chi-square statistics generated by the association tests and dividing it by the expected median of the Chi-square distribution, which is ~ 0.456. A lambda value greater than 1 indicates potential issues such as the presence of population substructure, cryptic relatedness, or systematic genotyping errors, indicating the requirement for the correction of the issue.

SNP annotations

The latest version of the *C. angulata* genome assembly (refSeq *ASM2576567V3*) does not yet have annotation in public databases. The annotation is only available for the previous build *ASM2561292V2*. In order to annotate the SNPs for their intersecting or closest genes, the SNPs were mapped against the ASM2561292V2 reference genome using 100 bases flanking sequence on either side of the SNPs (total sequence Length 201 bp). BLASTn program was used for mapping purpose. The estimated co-ordinate of the SNPs on the previous assembly was then used to identify the overlapping nearest genes using "Closest" option in Bedtools (version 2.29.2). Any genes within 10 kb up or downstream of the significant SNPs were also extracted using Bedtools "Intersect" option from gene annotation file for *C. angulata* from Ensembl Metazoa (Release 60). The Gene Ontology (GO) terms and their associated accession numbers were obtained from Ensembl Metazoa (Release 60) using Biomart.

Results

Trait data

The summary statistics of the different traits studied are presented in Table 1 and Fig. 1 boxplots, which show moderate to substantial variations for most traits. The Coefficient of

Table 1 Basic statistics of phenotypic data in the *Crassostrea angulata* population traits

Traits	Unit	*n*	Mean	SD	Min	Max	Coefficient of variation
Shell length	cm	487	7.40	1.33	3.50	11.40	18.00%
Shell width	cm	488	4.16	0.74	1.50	6.50	17.88%
Shell height	cm	356	2.77	0.55	1.10	5.80	19.95%
Shell weight	g	484	38.50	15.37	9.45	99.36	39.91%
Condition index	Index	281	67.56	21.65	14.29	99.93	32.05%
Shell colour a*	CIELab	302	2.33	3.70	−2.07	54.96	158.74%
Shell colour b*	CIELab	302	11.55	4.23	3.99	43.02	36.63%
Shell colour L*	CIELab	302	23.41	12.19	6.49	69.71	52.05%
Mantle colour a*	CIELab	443	4.38	3.43	0.16	24.60	78.19%
Mantle colour b*	CIELab	445	10.09	3.93	1.02	39.02	39.99%
Mantle colour L*	CIELab	445	24.20	8.73	7.24	63.70	36.06%

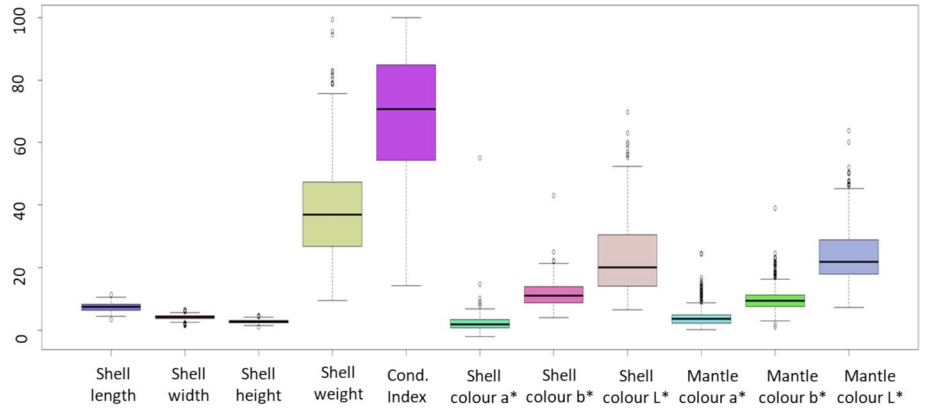

Fig. 1 Box plots showing the distribution of measured phenotypic trait values in Portuguese oysters. Each trait is presented in its respective unit

Variation (CV) values provided in Table 1 allow the comparison of the level of variation among traits measured on different scales. Within the growth-related traits, shell weight exhibited the largest variation around the mean (CV = 39.91%) followed by the condition index (32.05%). The lowest variation is observed in shell width (17.88%), followed by shell length (18.00%). In contrast, the shell and mantle colour-related traits show very large variations around the mean, with CV values ranging from 36.06% for mantle colour L* to 158.74% for shell colour a*. While the full extent of these colour variations may not be readily discernible to the naked eye, they were precisely and quantitatively detected by the calibrated CIELab colorimeter used in the present study, indicating substantial inherent biological diversity in shell pigmentation among individuals.

GWAS results

Assessing the Q-Q plots and genomic inflation factor

Uncorrected population structure in GWAS analyses can result in false positive signals. In our study, we employed mixed models that incorporate a genomic relationship matrix to correct for population structure and cryptic relationships. The first step in assessing the GWAS results is to examine the QQ plots and the genomic inflation factor, or λ values (Fig. 2). The QQ plots from our GWAS analyses indicate effective correction of population structure, as most of the markers lie around the diagonal line, reflecting a correspondence between the expected and observed *p*-values. While the QQ plot provides a visual representation of the distribution of observed *p*-values against the expected distribution under the null hypothesis of no association, the λ value quantifies the degree of inflation in the test statistics due to population structure or other confounding factors.

As shown in Fig. 2, the λ values for the traits varied between 0.84 and 1.16. While a λ value greater than 1 may indicate inflation in the test statistics due to population structure or cryptic relatedness, in practice, λ values up to 1.1 are generally considered acceptable (Williams et al. 2021). Conversely, λ values below 1 indicate deflated *p*-values, which may result from over-correction for population structure. For *shell colour a**, we observed a λ

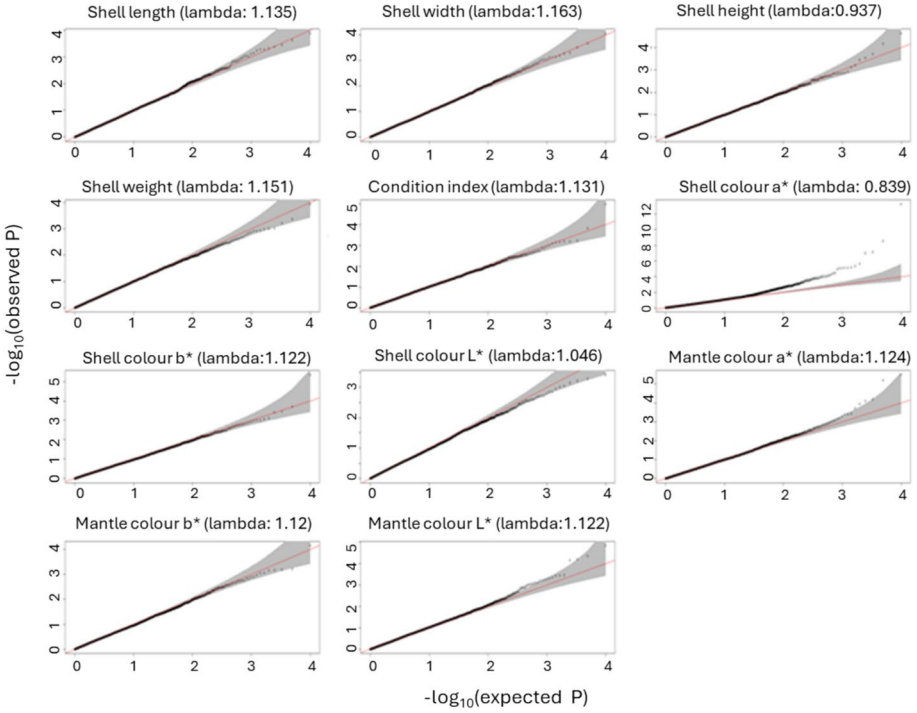

Fig. 2 Q-Q plots from GWAS analyses performed for the 11 traits. X-axis in each plot represents -log10(expected *p*-value) and Y axis represents -log10(observed *p*-value)

value of only 0.84, which is lower than the accepted threshold of 0.9, indicating possible deflation. To account for this, the *p*-values were adjusted by the estimated value of λ.

SNPs associated with growth and colour traits

The GWAS results identified 31 SNPs associated with the traits (Fig. 3, Table 2). Only four SNPs were found to be significantly associated with growth traits — one for shell width, two for height and one for condition index. In contrast, a larger number of significant associations ($n = 29$) were detected for colour traits. These included 4 SNPs for mantle colour a*, 1 for mantle colour b*, 4 for mantle colour L*, 19 for shell colour a*, and 1 for shell colour b* (Table 2). Two of these SNPs were found significantly associated with both mantle colour a* and b* (2:67804078_G_T) and shell colour a* and b* (3:2494231_T_A). Although a larger number of SNPs were identified for shell colour a*, results need to be treated with some caution as even after correcting for genomic inflation the model fit remained suboptimal (Fig. 2). The proportion of phenotypic variance explained by individual SNP is generally small for all traits ranging between 0.05% and 2.16% (average: 0.55%).

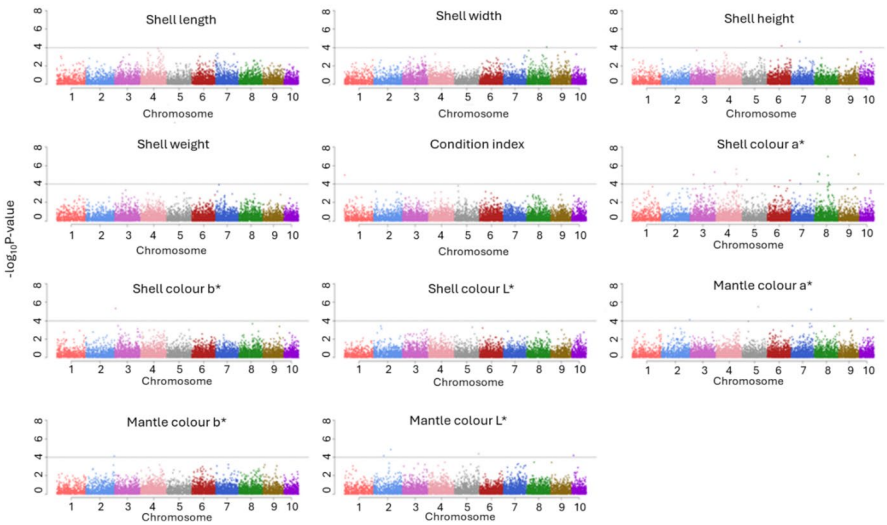

Fig. 3 Manhattan plots of GWAS results for the 11 traits

Candidate genes

Except for three SNPs, all other significant SNPs intersected with protein-coding genes or lncRNA or pseudogenes. Out of the 31 SNPs, 24 are intersected with protein coding genes (Table 2). For growth related traits, the intersected genes includes *LOC128167327* (WD repeat domain phosphoinositide-interacting protein 3-like or *WIPI3*-like) with potential roles in autophagy (Almannai et al. 2022), *LOC128177318* (centrosomal protein of 128 kDa-like or *CE128-like*) — involved in protein localization (GeneCard 2024) and *LOC128174230* which is an uncharacterized protein coding gene but has a number of associated GO terms including transcription regulation (GO:0140110 and GO:0006355) and molecular transducer activity (GO:0060089).

The genes intersecting with colour traits include glucose dehydrogenase and others homologous to Neurobeachin, *RalGPS1*, chymotrypsin-like serine proteinase, anoctamin-7, *DMSR-1, POP1*, fez family zinc finger protein erm, neo-calmodulin, *Elf-3* (transcription factor), *Rab-24*, myoferlin, mitochondrial carrier homolog 2, tafazzin, filamin, and *RAPH1*. Among these glucose dehydrogenases, neurobeachin-like and *POP1*-like explained relatively larger proportion of variance (1%−1.6%). *LOC128183296* (Neo-calmodulin-like) gene was found to intersect with two significant SNPs.

Since candidate genes may not necessarily be the ones closest to the significant SNPs and could be located further away, we searched for all genes within 10 kb upstream or downstream of all significant SNPs. This analysis identified a total of 60 genes (Supplementary Table 2). Among these, 21 genes have one or more associated Gene Ontology (GO) terms (Fig. 4). For growth-related traits only three more additional genes (that are not reported in Table 2) are detected in relation to shell height, including *LOC128166472* (enoyl-CoA hydratase, mitochondrial-like) involved in catalytic activity (GO:0003824) and *LOC128177340* (eukaryotic translation initiation factor 3 subunit K-like) with role in translation regulation (GO:0045182). For colour-related traits, the genes with

Table 2 Significant markers ($p < 0.0001$) associated with growth and colour-related traits identified using GBLUP-GWAS multi-loci mixed model methodology along with the overlapping or closest genes. Gene ontology (GO) terms and their accessions are extracted using the Ensembl Metazoa Biomart tool

Traits	*SNP*	*p-value*	*Proportion of variance*	*Gene* [#]	GO Slim terms and accession[$]
Shell width	8:47854918_G_C	8.72E-05	0.001124	LOC128174230 (protein coding; uncharacterized)	Transcription regulator activity (GO:0140110); molecular transducer activity (GO:0060089); DNA binding (GO:0003677); regulation of DNA-templated transcription (GO:0006355)
Shell height	6:34085837_T_A	6.95E-05	0.021595	LOC128177318 (centrosomal protein of 128 kDa-like or *CE128-like*)	
Shell height	7:20219456_T_G	2.34E-05	0.015801	LOC128167327 (WD repeat domain phosphoinositide-interacting protein 3-like or *WIPI3*-like)	
Condition index	1:602193_C_G	1.05E-05	0.006359	LOC128171427 (protein coding; uncharacterized)	
Shell colour a*	3:2494231_T_A	5.11E-14	0.001733	LOC128157526 (pseudogene)	
Shell colour a*	3:8244056_A_G	8.64E-06	0.010752	LOC128156661 (neurobeachin-like)	
Shell colour a*	3:34129196_G_T	9.29E-05	0.004193	LOC128157127 (ras-specific guanine nucleotide-releasing factor RalGPS1-like)	Molecular function regulator activity (GO:0098772)
Shell colour a*	3:58608947_C_A	4.77E-06	0.006099	LOC128156255 (chymotrypsin-like serine proteinase); ($d = 4663$ bp)	Catalytic activity (GO:0003824)
Shell colour a*	4:20211205_G_A	7.23E-05	0.004524	LOC128189452 (anoctamin-7-like)	
Shell colour a*	4:47752029_A_T	2.38E-06	0.00746	LOC128188897 (G-protein coupled receptor dmsr-1-like) ($d = 4394$ bp)	Molecular transducer activity (GO:0060089)
Shell colour a*	4:47991880_C_T	7.53E-06	0.004237	LOC128186993 (pseudogene)	
Shell colour a*	5:487872_A_G	2.86E-09	0.002946	LOC128160124 (protein coding; uncharacterized)	
Shell colour a*	5:11090119_T_A	3.29E-05	0.004213	LOC128179673 (lncRNA)	
Shell colour a*	6:53745921_T_A	3.78E-05	0.004524	LOC128177211 (protein coding; uncharacterized)	
Shell colour a*	7:22062737_G_A	9.45E-05	0.009659	LOC128164428 (ribonucleases P/MRP protein subunit POP1-like)	tRNA metabolic process (GO:0006399)
Shell colour a*	8:8201293_G_T	5.09E-05	0.006963	LOC128156130 (fez family zinc finger protein erm-like)	
Shell colour a*	8:11478285_A_T	8.60E-06	0.004642	LOC128183296 (neo-calmodulin-like)	

Table 2 (continued)

Traits	*SNP*	*p-value*	*Proportion of variance*	*Gene* [#]	GO Slim terms and accession[$]
Shell colour a*	8:11478308_C_T	6.85E-06	0.005432	LOC128183296 (neo-calmodulin-like)	
Shell colour a*	8:32561926_C_T	1.07E-05	0.003876	LOC128190074 (protein coding; uncharacterized)	
Shell colour a*	8:32561937_G_A	9.75E-08	0.00374	LOC128190074 (protein coding; uncharacterized)	
Shell colour a*	8:34338642_A_G	6.60E-05	0.006025	LOC128173812 (ETS-related transcription factor Elf-3-like)	Regulation of DNA-templated transcription (GO:0006355); transcription regulator activity (GO:0140110); DNA binding (GO:0003677)
Shell colour a*	9:39305803_T_G	7.15E-08	0.000573	LOC128179414 (ras-related protein Rab-24-like)	Catalytic activity (GO:0003824); hydrolase activity (GO:0016787); GTPase activity (GO:0003924)
Shell colour a*	9:47971335_G_T	7.71E-06	0.004452	LOC128181377 (myoferlin-like)	
Shell colour b*	3:2494231_T_A	4.55E-06	0.001925	LOC128157526 (pseudogene)	
Mantle colour a*	2:67804078_G_T	7.75E-05	0.005189	LOC128183915 (mitochondrial carrier homolog 2-like)	
Mantle colour a*	5:38218066_T_C	3.24E-06	0.002341	LOC128161544 (tafazzin-like)	Lipid metabolic process (GO:0006629); transferase activity (GO:0016740); catalytic activity (GO:0003824)
Mantle colour a*	7:47947761_T_A	6.26E-06	0.002337	LOC128180625 (pseudogene)	
Mantle colour a*	9:29451381_G_A	6.54E-05	0.001069	LOC128182358 (filamin-C-like)	Cytoskeletal protein binding (GO:0008092); cytoskeleton organization (GO:0007010)
Mantle colour b*	2:67804078_G_T	7.09E-05	0.004962	LOC128183915 (mitochondrial carrier homolog 2-like)	
Mantle colour L*	2:24261515_A_C	7.06E-05	0.001951	LOC128185182 (pseudogene)	
Mantle colour L*	2:41141537_C_T	1.43E-05	0.003521	LOC128185328 (pseudogene)	
Mantle colour L*	5:57566637_A_G	4.36E-05	0.016359	LOC128184820 (glucose dehydrogenase)	catalytic activity (GO:0003824); oxidoreductase activity (GO:0016491)
Mantle colour L*	10:4043042_C_G	6.23E-05	0.002296	LOC128163538 (ras-associated and pleckstrin homology domains-containing protein 1 (RAPH1) -like)	

[#] *For SNPs that did not intersect with any gene, the closest gene is reported, and the distance (d) between the SNP and the gene is provided. GO Slim terms and accessions are obtained from Ensembl Metazoa (release 60) using BioMart.*

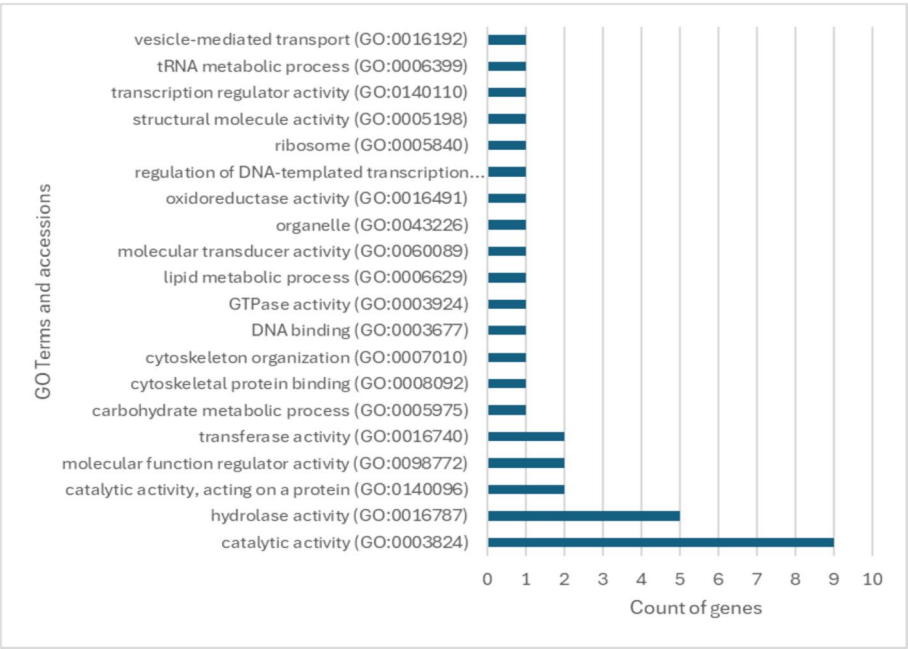

Fig. 4 Gene Ontology (GO) terms associated with candidate genes located within 10 kb upstream or downstream of significant SNPs for colour-related traits

available GO terms show diverse roles but the majority are associated with catalytic activities (GO:0003824, GO:0140096) and hydrolase activities (GO:0016787) (Fig. 4).

Discussion

The present study provides novel insights into the genetic architecture of key commercial traits in the Portuguese oyster (*C. angulata*), focusing on shell growth, condition index, and shell and mantle colour. This is the first study to perform GWAS on colour traits in this species — an important commercial attribute with potential for selective improvement. By identifying SNPs significantly associated with these traits and closely linked to relevant candidate genes, our study contributes valuable knowledge and resources to support the selective breeding of *C. angulata*.

Our study utilized phenotypic and genotypic data from individuals belonging to a *C. angulata* commercial aquaculture population. The samples were collected from two consecutive generations, encompassing approximately 90 full-sib families (57 families from the first generation and 33 from the second generation). Although several approaches are available for identifying candidate genes or genetic variants associated with phenotypic traits — most notably QTL mapping and GWAS — we opted for the latter approach in this study. While the family-based sample structure could also support QTL mapping, which relies on within-family segregation of SNP genotypes with traits, this method typically requires large family sizes (100–200 individuals per family), rather than a large number of families with a few progenies, to achieve adequate

statistical power (e.g. see Massault et al. 2008). In our case, although we had large number of families, the family sizes ranged from only 12 to 15 individuals, limiting the power for effective QTL detection. In contrast, GWAS leverages linkage disequilibrium (LD) across the broader population (Visscher et al. 2017) — in our case, across all families from two generations, thereby providing greater power to detect associations that segregate beyond individual families. We therefore applied a GWAS approach to make full use of the available samples. We employed a GBLUP-based GWAS analysis, which uses the genomic relationship matrix to account for relatedness and population structure, offering a robust framework for association analysis. Compared to traditional GWAS methods that assess individual SNP effects, GBLUP considers the cumulative contribution of all SNPs across the genome, making it especially appropriate for analysing complex polygenic traits (Legarra et al. 2018; Wijesena et al. 2023). Its ability to simultaneously perform genomic prediction and association mapping further enhances its applicability to breeding programs, supporting the rationale behind our methodological choice.

Nevertheless, our study has certain limitations that should be acknowledged. First, the sample size was relatively small, reducing the statistical power of the GWAS analyses. Although 647 samples were initially collected, the number included in the final analyses ranged from 302 to 488 across different traits (Table 1). For complex traits, GWAS analyses typically require sample sizes of over 1000 individuals to achieve robust statistical power (Visscher et al. 2017). Second, the number of SNPs used was modest ($n = 9846$), limited by the resolution of the DArTseq genotyping platform. These limitations reflect the early stage of the Portuguese oyster breeding program in Vietnam and the resource constraints associated with it. Despite these challenges, our GWAS identified candidate SNPs near genes with plausible functional relevance, which may reflect a substantial level of LD in the studied population — commonly observed in farmed stocks derived from a limited number of founder individuals (Qanbary 2020). There are instances of other studies which have successfully identified meaningful associations using similarly limited datasets (e.g. see Wang et al. (2025) study on pearl mussel). Future work using larger populations and higher-density genotyping will be important to validate these findings and improve the resolution of trait–genotype associations.

Our study identified several candidate genes associated with the analysed traits. However, the effect sizes of the associated SNPs were relatively small (ranging from 0.05% to 2.16%), suggesting a polygenic architecture. This is consistent with previous findings in other species, where growth and pigmentation traits were found to be complex and polygenic (e.g. Ju and Mathieson 2021; Liu et al. 2024; Wang et al. 2025). Given the limited understanding of trait architecture in farmed bivalves, such insights are valuable for developing more effective breeding programs. Specifically, the polygenic nature of these traits implies that genomic selection is likely to be more effective than marker-assisted selection, which is better suited for traits controlled by major genes. Genomic selection, on the contrary, leverages genome-wide SNP data to capture the combined effects of many loci, making it more appropriate for complex polygenic traits. The significant SNPs identified in our study provide biologically informative markers that could be incorporated into SNP panels to enhance genomic prediction. Although their individual effects are small, including GWAS-identified SNPs in genomic selection models has been shown to improve prediction accuracy (Meuwissen et al., 2024). For instance, Meuwissen et al., (2024) developed the GWABLUP method, which assigns greater weight to SNPs identified through GWAS, Leading to a 10–13% improvement in prediction reliability over standard GBLUP.

Some of the notable candidate genes detected in our study include the *centrosomal protein of 128 kDa-like* (*CE128*-like) and the *WD repeat domain phosphoinositide-interacting protein 3-like* (*WIPI3*-like) associated with growth-related trait — shell height. These genes are implicated in protein localization and autophagy, respectively, processes that are critical for cellular growth and maintenance (Almannai et al. 2022). Additionally, several yet uncharacterized genes were detected in relation to shell width and condition index, warranting further investigation. In relation to colour-related traits, a variety of candidate genes were detected, many of which are associated with GO terms related to catalytic and hydrolase activities. The expression of colour traits in molluscs is inherently complex, as it is influenced by the interplay of genetic, dietary, and environmental factors (Saenko & Schilthuizen, 2020). This complexity may explain the small effect sizes of the significant SNPs identified in our study. It could also account for the limited overlap between the genes identified in our research and those reported in other studies on colour traits in oysters and molluscs in general. Only one candidate gene — *LOC128183296* (*Neo-calmodulin-like*) — found in our study in relation to shell colour is known to participate in calcium-binding processes critical for pigmentation in mollusc species (Saenko & Schilthuizen, 2020; Wang et al. 2022; Hu et al. 2025). In a previous study, the calmodulin gene was found to be associated with shell colour in pearl oysters (Liu et al. 2023). The scarcity of genetic research on colour traits in oysters may also explain the limited confirmation of our results from the literature. Some other notable colour-associated SNPs detected in the present study (with effect size > 1%) were linked to genes, *glucose dehydrogenase*, *Neurobeachin*, and *POP1-like*. Although the role of these genes in oyster pigmentation will need to be elucidated through further studies, certain interpretation can be proposed based on known functions of the genes. For instance, *glucose dehydrogenase* (*G6PD*) is an enzyme primarily involved in the pentose phosphate pathway, a carbohydrate metabolic process that generates NADPH, a critical reducing agent used in various cellular activities (GeneCard). NADPH has been shown to regulate melanin synthesis in humans (Yamaguchi et al. 2010). Furthermore, alloenzymes of *G6PD* have been associated with pigmentation in the bacterium *Serratia marcescens* (Gargallo et al. 1987). These findings suggest that the *G6PD* gene could play a role in pigmentation processes across different organisms, making it a strong candidate for colour-related traits in oysters. Similarly, the *Neurobeachin* gene primarily functions in neuronal processes and is involved in membrane trafficking (Medrihan et al. 2009). This role could influence the secretion of proteins or pigments from mantle cells to the shell as in bivalves, shell colouration often depends on pigments produced by mantle tissues (Saenko & Schilthuizen, 2020). On the other hand, the *POP1* gene, which plays a role in tRNA metabolism, might influence shell colouration indirectly. This could occur through the regulation of RNA processing and the translation of pigment-related proteins or by affecting the expression of biomineralization-related genes via modulation of RNA stability (e.g. see Parvizi et al. 2023). While further studies are needed to confirm their roles, these candidate genes and their 31 associated SNPs provide valuable markers for potential application in genomic selection for pigmentation traits in oysters.

In this study, we employed the DArTseq genotype-by-sequencing approach to generate genome-wide SNP markers. While DArTseq is a valuable and cost-effective method for simultaneously detecting and genotyping markers, routine implementation of genomic selection in breeding programs would require high-density SNP genotyping arrays. These arrays enable high-throughput genotyping of hundreds or thousands of individuals, making them more suitable for large-scale selective breeding. Although the development and implementation of high-density SNP arrays have cost implications, genomic selection

offers significant advantages, including higher prediction accuracy and the ability to maintain genetic diversity, thereby supporting the sustainability of breeding efforts. The potential for applying genomic selection in oyster species and its superiority over pedigree-based selection has been demonstrated in studies on *C. gigas* (Jourdan et al. 2023). Additionally, research on *C. angulata* has shown that the cost of genomic selection can be mitigated by using low-density SNP panels for genotyping, followed by imputation in offspring generations (Vu et al. 2021a, b, c). Therefore, it will be essential to generate additional markers through whole-genome resequencing projects in the future and to develop both high- and low-density SNP genotyping arrays. Additionally, building a comprehensive reference database of genotypes for imputation will be crucial for optimizing genomic selection strategies in oyster breeding programs.

Conclusions

In conclusion, this study advances our understanding of the genetic architecture underlying growth and pigmentation traits in the Portuguese oyster (*Crassostrea angulata*). Using GBLUP-based GWAS, we identified several SNPs and candidate genes significantly associated with these traits, despite their small effect sizes—suggesting the polygenic control of the traits. This suggests that genomic selection would be a more suitable breeding strategy than marker-assisted selection. However, successful implementation will require developing SNP genotyping arrays, along with robust high-density SNP genotype databases to enable cost-effective imputation. Our findings provide a valuable foundation for genomic selection in *C. angulata*, offering initial candidate markers and a resource for SNP panel development. While challenges remain in translating these insights into applied breeding, this work lays important groundwork for improving the efficiency and profitability of *C. angulata* aquaculture through genomics-informed approaches.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s10499-025-02210-6.

Acknowledgements The authors would like to thank Professor Ziqiang Han of Jimei University, China, for his valuable comments on this manuscript, as well as Ms. Kieu Thi Hong and Ms. Phan Thi Duc Hien (undergraduate students) from the Faculty of Biotechnology, Hanoi University of Pharmacy, Hanoi, Vietnam, and Ms. Tran Luu Ngoc Ha (undergraduate student) from the University of Science, Vietnam National University, Hanoi, Vietnam, for their assistance. In addition, we thank Dr Cao Truong Giang and Ms Tran Thi Nguyet Minh at Research Institute for Aquaculture number 1, Cat Ba, Hai Phong, Vietnam for their assistance.

Author contribution Sang Van Vu and Wayne O'Connor were involved in the conceptualization of the study and funding acquisition. Sang Van Vu, Cedric Gondro, and Almas A. Gheyas were involved in formal analyses of the data and preparing the first draft. All other authors (In Van Vu, Thu Thi Anh Nguyen, Hsu Htoo, Shantanu Kundu, Kim Hyun Woo, Soo Rin Lee, Tran Dang Khanh, Tiep Khac Nguyen, and Hien Van Doan) contributed to different aspects of the analyses and in reviewing and editing the draft manuscript. All authors read and approved the final manuscript.

Funding This study was funded by the Australian Centre for International Agricultural Research, Australia. In addition, this work is supported by the Chey Institute for Advanced Studies' International Scholar Exchange Fellowship for the academic year of 2024–2025.

Data availability The data used in this work are available from the corresponding author upon reasonable request.

Declarations

Competing interests The authors declare no competing interests.

References

Ajithkumar M, D'ambrosio J, Travers MA, Morvezen R, Degremont L (2025) Genomic selection for resistance to one pathogenic strain of *Vibrio splendidus* in blue mussel *Mytilus edulis*. Front Genet 15:1487807

Almannai M, Marafi D, El-Hattab AW (2022) WIPI proteins: biological functions and related syndromes. Front Mol Neurosci 15:1011918

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. J Mol Biol 215:403–410

Brigida D (2023) Oysters: an unsung hero in a changing climate. Available at: https://www.worldwildlife.org/stories/oysters-an-unsung-hero-in-a-changing-climate#:~:text=Oysters%20possess%20a%20remarkable%20ability,carbon%20dioxide%20from%20the%20atmosphere. Accessed 06 Feb 2025

Divilov K, Schoolfield B, Morga B, Dégremont L, Burge CA, Mancilla Cortez D, Friedman CS, Fleener GB, Dumbauld BR, Langdon C (2019) First evaluation of resistance to both a California OsHV-1 variant and a French OsHV-1 microvariant in Pacific oysters. BMC Genet 20:1–9

Fardo DW, Becker KD, Bertram L, Tanzi RE, Lange C (2009) Recovering unused information in genome-wide association studies: the benefit of analyzing SNPs out of Hardy-Weinberg equilibrium. Eur J Hum Genet 17(12):1676–1682. https://doi.org/10.1038/ejhg.2009.85

Gargallo D, Lorén JG, Guinea J, Viñas M (1987) Glucose-6-phosphate dehydrogenase alloenzymes and their relationship to pigmentation in *Serratia marcescens.* Appl Environ Microbiol 53(8):1983-1986. https://doi.org/10.1128/aem.53.8.1983-1986.1987

GeneCard (2024) CEP128 Gene - Centrosomal Protein 128. Available at: https://www.genecards.org/cgi-bin/carddisp.pl?gene=CEP128. Accessed 15 Jan 2025

Gogarten SM, Bhangale T, Conomos MP, Laurie CA, McHugh CP, Painter I, Zheng X, Crosslin DR, Levine D, Lumley T, Nelson SC, Rice K, Shen J, Swarnkar R, Weir BS, Laurie CC (2012) GWASTools: an R/Bioconductor package for quality control and analysis of genome-wide association studies. Bioinformatics 28(24):3329–3331. https://doi.org/10.1093/bioinformatics/bts610

He X, Li C, Qi H, Meng J, Wang W, Wu F, Li L, Zhang G (2021) A genome-wide association study to identify the genes associated with shell growth and shape-related traits in *Crassostrea gigas*. Aquaculture 543:736926 (article ID: 736926)

Houston RD, Bean TP, Macqueen DJ, Gundappa MK, Jin YH, Jenkins TL, Selly SLC, Martin SAM, Stevens JR, Santos EM, Davie A, Robledo D (2020) Harnessing genomics to fast-track genetic improvement in aquaculture. Nature Review Genetics 21:389–409. https://doi.org/10.1038/s41576-020-0227-y

Hu B, Yu H, Xu C, Kong L, Liu S, Li Q (2025) Shell colour diversity in marine molluscs: from current knowledge to future aquaculture applications. Rev Aquacult 17(3):e70038

Jourdan A, Morvezen R, Enez F, Haffray P, Lange A, Vétois E, Allal F, Phocas F, Bugeon J, Dégremont L, Boudry P (2023) Potential of genomic selection for growth, meat content and colour traits in mixed-family breeding designs for the Pacific oyster *Crassostrea gigas*. Aquaculture 576:739878. https://doi.org/10.1016/j.aquaculture.2023.739878

Ju D, Mathieson I (2021) The evolution of skin pigmentation-associated variation in West Eurasia. Proc Natl Acad Sci U S A 118(1):e2009227118. https://doi.org/10.1073/pnas.2009227118

Kahn BE, Wansink B (2004) The influence of assortment structure on perceived variety and consumption quantities. J Consum Res 30:519–533

Kause A, Quinton C, Airaksinen S, Ruohonen K, Koskela J (2011) Quality and production trait genetics of farmed European whitefish, *Coregonus lavaretus*. J Anim Sci 89:959–971

Kilian A, Wenzl P, Huttner E, Carling J, Xia L, Blois H, Caig V, Heller-Uszynska K, Jaccoud D, Hopper C (2012) Diversity arrays technology: a generic genome profiling technology on open platforms. In: Pompanon F, Bonin A (eds) Data Production and Analysis in Population Genomics. Methods in Molecular Biology, vol 888. Humana Press, Totowa, NJ. https://doi.org/10.1007/978-1-61779-870-2_5

Lawrence DR, Scott GI (1982) The determination and use of condition index of oysters. Estuaries 5(1):23–27. https://doi.org/10.2307/1352213

Le TNP, Vu SV, Ugalde SC, Subramanian S, Gilmour A, Dove M, Vu IV, Geist J, Tran TNT, Gondro C, Cao GT, Le TT, Nguyen TM, Ngo TKN, Vu TTH, H. K. A. P, Knibb W, O'Connor W (2023) The genetics and breeding of the Portuguese oyster, *Crassostrea angulata*: lessons, experiences, and challenges in Vietnam. Front Mar Sci 10:1161009

Legarra A, Ricard A, Varona L (2018) GWAS by GBLUP: single and multimarker EMMAX and Bayes factors, with an example in detection of a major gene for horse gait. G3 Genes|genomes|genetics 8(7):2301–2308. https://doi.org/10.1534/g3.118.200336

León K, Mery D, Pedreschi F, León J (2006) Color measurement in Lab* units from RGB digital images. Food Research International 39(10):1084–1091. https://doi.org/10.1016/j.foodres.2006.03.006

Liu Y, Wang Z, Guo C, Li S, Li Y, Huang R, Deng Y (2023) Transcriptome and exosome proteome analyses provide insights into the mantle exosome involved in nacre colour formation of pearl oyster *Pinctada fucata martensii*. Comp Biochem Physiol D Genomics Proteomics 48:101151. https://doi.org/10.1016/j.cbd.2023.101151

Liu J, Yin M, Ye Z, Hu J, Bao Z (2024) Harnessing hue: advances and applications of fish skin pigmentation genetics in aquaculture. Fishes 9:220. https://doi.org/10.3390/fishes9060220

Loh PR, Danecek P, Palamara PF, Fuchsberger C, A Reshef Y, K Finucane H, Schoenherr S, Forer L, McCarthy S, Abecasis GRR (2016) Reference-based phasing using the Haplotype Reference Consortium panel. Nat Genet 48:1443–1448

Massault C, Bovenhuis H, Haley C, de Koning D-J (2008) QTL mapping designs for aquaculture. Aquaculture 285(1–4):23–29. https://doi.org/10.1016/j.aquaculture.2008.06.040

Medrihan L, Rohlmann A, Fairless R, Andrae J, Döring M, Missler M, Zhang W, Kilimann MW (2009) Neurobeachin, a protein implicated in membrane protein traffic and autism, is required for the formation and functioning of central synapses. J Physiol 587(21):5095–5106. https://doi.org/10.1113/jphysiol.2009.178236

Meuwissen T, Eikje LS, Gjuvsland AB (2024) GWABLUP: genome-wide association assisted best linear unbiased prediction of genetic values. Genet Sel Evol 56:17. https://doi.org/10.1186/s12711-024-00881-y

NOAA Fisheries (2022) Global study sheds light on the valuable benefits of shellfish and seaweed aquaculture. Available at: https://www.fisheries.noaa.gov/feature-story/global-study-sheds-light-valuable-benefits-shellfish-and-seaweed-aquaculture (Accessed on 15 Jan 2025).

O'Connor W, Dove M, O'Connor S, Van In V, Lien M, TN VP (2019) Enhancing bivalve production in northern Vietnam & NSW. ACIAR Canberra, Australia. Available at: https://www.aciar.gov.au/publication/technical-publications/enhancing-bivalve-production-northern-vietnam-and-australia-final-report. Accessed on 15 Jan 2025

Parvizi F, Akbarzadeh A, Farhadi A, Arnaud-Haond S, Ranjbar MS (2023) Expression pattern of genes involved in biomineralization in black and orange mantle tissues of pearl oyster, *Pinctada persica*. Front Mar Sci. https://doi.org/10.3389/fmars.2022.1038692

Peñaloza C, Barria A, Papadopoulou A, Hooper C, Preston J, Green M, Helmer L, Kean-Hammerson J, Nascimento-Schulze JC, Minardi D (2022) Genome-wide association and genomic prediction of growth traits in the European flat oyster (*Ostrea edulis*). Front Genet 13:926638

Qanbari S (2020) On the extent of linkage disequilibrium in the genome of farm animals. Front Genet 10:1304. https://doi.org/10.3389/fgene.2019.01304

Saenko SV, Schilthuizen M (2021) Evo-devo of shell colour in gastropods and bivalves. Curr Opin Genet Dev 69:1–5. https://doi.org/10.1016/j.gde.2020.11.009

Seafish (2020) Oysters. Available at: https://www.seafish.org/responsible-sourcing/aquaculture-farming-seafood/species-farmed-in-aquaculture/aquaculture-profiles/oysters/. Accessed 06 Feb 2025

Ugalde SC, Vu SV, Giang CT, Ngoc NTH, Tran TKA, Mullen JD et al (2023) Status, supply chain, challenges, and opportunities to advance oyster aquaculture in northern Vietnam. Aquaculture 572:739548. https://doi.org/10.1016/j.aquaculture.2023.739548

Van Raden PM (2008) Efficient methods to compute genomic predictions. J Dairy Sci 91:4414–4423

Vasimuddin M, Misra S, Li H, Aluru S (2019) Efficient architecture-aware acceleration of BWA-MEM for multicore systems. In: 2019 IEEE international parallel and distributed processing symposium (IPDPS). IEEE, pp 314–324

Visscher PM, Wray NR, Zhang Q, Sklar P, McCarthy MI, Brown MA, Yang J (2017) 10 years of GWAS discovery: biology, function, and translation. Am J Hum Genet 101(1):5–22. https://doi.org/10.1016/j.ajhg.2017.06.005

Vu SV, Knibb W, Nguyen NT, Vu IV, O'connor W, Dove M, Nguyen NH (2020) First breeding program of the Portuguese oyster *Crassostrea angulata* demonstrated significant selection response in traits of economic importance. Aquaculture 518:734664

Vu SV, Gondro C, Nguyen NT, Gilmour AR, Tearle R, Knibb W, Dove M, Vu IV, Khuong LD, O'Connor W (2021) Prediction accuracies of genomic selection for nine commercially important traits in the Portuguese oyster (*Crassostrea angulata*) using DArT-seq technology. Genes 12(2):210

Vu SV, Knibb W, Gondro C, Subramanian S, Nguyen NTH, Alam M, Dove M, Gilmour AR, Vu IV, Bhyan S, Tearle R, Khuong LD, Le TS, O'Connor W (2021) Genomic prediction for whole weight, body shape, meat yield, and colour traits in the Portuguese oyster *Crassostrea angulata*. Front Genet 12:661276. https://doi.org/10.3389/fgene.2021.661276

Vu SV, Premachandra H, O'Connor W, Nguyen NT, Dove M, Van Vu I, Le TS, Vendrami DL, Knibb W (2021) Development of SNP parentage assignment in the Portuguese oyster *Crassostrea angulata*. Aquac Rep 19:100615

Vu SV, Kumar M, Rastas P et al (2024) High-density linkage map and single nucleotide polymorphism association with whole weight, meat yield, and shell shape in the Portuguese oyster, *Crassostrea angulata*. Aquacult Int 32(1):10109–10122. https://doi.org/10.1007/s10499-024-01652-8

Wang Z, Liu Y, Zheng Z, Deng Y (2022) A genome-wide association study identifies SNPs and candidate genes associated with nacre colour of pearl oyster *Pinctada fucata martensii*. Aquaculture 561:738–743

Wang Z, Hu H, Wang H, Yan L, Zhang Y, Wang H, Lv X, Li J, Bai Z (2025) Genome-wide association analysis reveals genetic architecture of growth and inner shell color traits in freshwater pearl mussel *Hyriopsis cumingii* (Lea 1852). Aquaculture 595(2):741658

Wijesena HR, Nonneman DJ, Snelling WM, Rohrer GA, Keel BN, Lents CA (2023) gBLUP-GWAS identifies candidate genes, signaling pathways, and putative functional polymorphisms for age at puberty in gilts. J Anim Sci 101:skad063. https://doi.org/10.1093/jas/skad063

Williams CJ, Li Z, Harvey N et al (2021) Genome wide association study of response to interval and continuous exercise training: the Predict-HIIT study. J Biomed Sci 28(1):37. https://doi.org/10.1186/s12929-021-00733-7

Woolmer A (n.d.) The nutritional benefits of shellfish. Available at: https://shellfish.org.uk/wp-content/uploads/2023/11/23239Nutritional_benefits_of_shellfish1.pdf. Accessed 06 Feb 2025

Wu L, Gong S, Li H, Ke C, Shi B (2023) Genome-wide association study and candidate gene analysis of cadmium accumulation in Fujian oysters (*Crasosstrea angulata*). Aquaculture 572:739546

Xie J, Ning Y, Han Y, Su C, Zhou X, Wu Q, Guo X, Qi J, Ge H, Ke Y et al (2024) Identification of SNPs and candidate genes associated with growth using GWAS and transcriptome analysis in Portuguese oyster (*Magallana angulata*). Fishes 9(12):471. https://doi.org/10.3390/fishes9120471

Xing D, Li Q, Kong L, 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

Yamaguchi Y, Hearing V, Maeda A, Morita A (2010) NADPH: quinone oxidoreductase-1 as a new regulatory enzyme that increases melanin synthesis. J Invest Dermatol 130(3):645–647. https://doi.org/10.1038/jid.2009.378

Yu F, Peng W, Tang B, Zhang Y, Wang Y, Gan Y, Luo X, You W, Gwo J-C, Chen N (2021) A genome-wide association study of heat tolerance in Pacific abalone based on genome resequencing. Aquaculture 536:736436

Zhao M, Thaimuangphol W, Hong Y, Yan Z, Chen Z, Jin M, Zheng A, Wang B, Wang Z (2023) Identification of growth-related SNPs and genes in the genome of the pearl oyster (*Pinctada fucata*) using GWAS. Fishes 8:296

Zhu X, Zhang J, Hou X, Liu P, Lv J, Xing Q, Huang X, Hu J, Bao Z (2021) A genome-wide association study identifies candidate genes associated with shell colour in bay scallop *Argopecten irradians irradians*. Front Mar Sci 8:742330

Authors and Affiliations

Sang Van Vu[1] · Wayne O'Connor[2] · In Van Vu[3] · Cedric Gondro[4] · Thu Thi Anh Nguyen[5] · Shantanu Kundu[6,7] · Kim Hyun Woo[8,9] · Soo Rin Lee[8,10] · Tran Dang Khanh[11] · Tiep Khac Nguyen[12] · Hien Van Doan[13,14] · Hsu Htoo[8] · Almas A. Gheyas[15]

✉ Sang Van Vu
sangvv@vnu.edu.vn

✉ Almas A. Gheyas
almas.gheyas@stir.ac.uk

1 Faculty of Biology, University of Science, Vietnam National University, Hanoi, Vietnam

2 NSW Department of Primary Industries, Port Stephens Fisheries Institute, Taylors Beach, NSW 2316, Australia

3 Faculty of Advanced Technology, Vietnam Japan University, Vietnam National University, Hanoi, Vietnam

4 Department of Animal Science, College of Agriculture and Natural Resources, Michigan State University, East Lansing, MI 48824, USA

5 Institute for Biotechnology and Environment, Nha Trang University, Khanh Hoa, Vietnam

6 Ocean and Fisheries Development International Cooperation Institute, College of Fisheries Science, Pukyong National University, Busan 48513, Republic of Korea

7 International Graduate Program of Fisheries Science, Pukyong National University, Busan 48513, Republic of Korea

8 Department of Marine Biology, Pukyong National University, Busan 48513, Republic of Korea

9 Marine Integrated Biomedical Technology Center, National Key Research Institutes in Universities, Pukyong National University, Busan 48513, Republic of Korea

10 Institute of Marine Living Modified Organisms, Pukyong National University, Busan 48513, Republic of Korea

11 Department of Nematology, University of California, Riverside, CA 92521, USA

12 Faculty of Biotechnology, Hanoi University of Pharmacy, Hanoi, Vietnam

13 Department of Animal and Aquatic Sciences, Faculty of Agriculture, Chiang Mai University, Chiang Mai 50200, Thailand

14 Functional Feed Innovation Center (FuncFeed), Faculty of Agriculture, Chiang Mai University, Chiang Mai 50200, Thailand

15 Institute of Aquaculture, University of Stirling, Stirling, Scotland, UK