Aquaculture International
December 2024, Volume 32 Issue 7 Pages 10109-10122
https://doi.org/10.1007/s10499-024-01652-8
https://archimer.ifremer.fr/doc/00907/101910/

High-density linkage map and single nucleotide polymorphism association with whole weight, meat yield, and shell shape in the Portuguese oyster, *Crassostrea angulata*

Van Vu Sang [1, *], Kumar Manoharan [2], Rastas Pasi [3], Boudry Pierre [4], Gheyas Almas [5], Bean Tim P. [6], Nguyen Mai Thi [7], Tran Khanh Dang [8], Geist Juergen [9], Nguyen Hoang Huy [10], O'connor Wayne [11], Tran Ha Luu Ngoc [1], Le Thang Toan [1], Cao Giang Truong [12], Nguyen Thu Thi Anh [13], Van Vu In [14]

[1] Faculty of Biology, VNU University of Science, Vietnam National University, Hanoi, Vietnam
[2] Australian Institute of Tropical Health and Medicine and Centre for Tropical Bioinformatics and Molecular Biology, James Cook University, Smithfield, QLD, 4878, Australia
[3] Institute of Biotechnology, University of Helsinki, Helsinki, Finland
[4] Département Ressources Biologiques Et Environnement, Ifremer, Plouzané, 29280, France
[5] Institute of Aquaculture, University of Stirling, Scotland, UK
[6] Roslin Institute, University of Edinburgh, Easter Bush Campus, Roslin, EH46 7AF, UK
[7] Faculty of Fisheries, Vietnam National University of Agriculture, Hanoi, Vietnam
[8] Vietnam National University of Agriculture, Hanoi, Vietnam
[9] Aquatic Systems Biology Unit, TUM School of Life Sciences, Technical University of Munich, Freising, Germany
[10] Institute of Genome Research, Vietnam Academy of Science and Technology, Hanoi, Vietnam
[11] NSW Department of Primary Industries, Port Stephens Fisheries Institute, Taylors Beach, NSW, 2316, Australia
[12] Research Institute for Aquaculture, Number 1, Bac Ninh, Vietnam
[13] Institute for Biotechnology and Environment, Nha Trang University, Khanh Hoa, Vietnam
[14] Biosecurity Unit, Faculty of Advanced Technologies and Engineering, Vietnam Japan University, Vietnam National University, Hanoi, Vietnam

* Corresponding author : Sang Van Vu, email address : sangvv@vnu.edu.vn

Abstract :

This study elucidates to generate a linkage map for Portuguese oysters, Crassostrea angulata, and recognise potential markers linked to economical traits in C. angulata. Genome-wide association studies (GWAS) using 19,475 single nucleotide polymorphisms (SNPs) generated by a genotyping-by-sequencing approach consisting of 647 individuals in two selection generations of C. angulata, consisting of 57 full-sib family of the first generation and 33 full-sib family of the second generation. GWAS identified six significant SNPs linked to whole weight ($P < 10-4$). These six SNPs explained 10.2% of the total genetic variance of whole weight. Additionally, three SNPs were identified linked to meat yield (i.e., soft tissue weight) and contributed to 9.2% of the total genetic variance for this trait. By linkage analysis on

the raw data, a total of 19,475 SNP markers were mapped to 10 linkage groups, with an average density of 19.1 SNPs per cM. This linkage map was developed and potential markers identified for whole weight and soft tissue weight provide a novel resource to characterise the genetic structure of economic traits in C. angulata.

Keywords : Crassostrea angulata, Linkage mapping, Genome-wide association study, Traits

Introduction

The Portuguese oyster (*Crassostrea angulata*) is a economic mollusc species in Asia, particularly Vietnam, where annual production has exceeded 50,000 tonnes in 2018 (O'Connor et al., 2019). Among oyster species being cultured worldwide, *Crassostrea gigas* and *C. angulata* are two dominant oyster species around the world and only differentiated by molecular genetic tools (Boudry et al., 1998; In et al., 2017), and they are able to hybridise where they coexist (Huvet et al., 2004, Lapègue et al., 2020) and generate progeny with spawning capacity (Huvet et al., 2002). Recently, the commercial importance of *C. angulata* has attracted considerable research attention on its economic traits and breeding improvement. A number of studies is increasingly using genomic information for oysters, such as evaluation of genetic diversity (Lapègue et al., 2020; Vu et al., 2017a), gene expression analysis of diploid and triploid oysters in relation to growth rate traits (Zeng et al., 2019), genomic selection (Vu et al., 2021a), parentage assignment (Vu et al., 2021b), and comparative chromosome-level genome assembly with the closely-related species *C. gigas* (Gagnaire et al., 2018; Qi et al., 2021; Qi et al., 2023, Wang et al., 2016, Wang et al., 2021). However, there have not yet been reports on identifying potential markers linked to traits of economic importance for this species.

Marker-assisted selection (MAS) can provide high efficiency, rapid implementation and simple application when compared to time-consuming and labour-intensive selection approaches applied in conventional selection such as increased accuracy, faster development of desired traits, and reduced breeding cycles. However, the efficacy of molecular selection depends on the determination of effective loci through the dissection of genetic linkage with techniques such as QTL mapping and GWAS. In combination with high density sequencing data, GWAS is able to identify trait-associated genes and genetic polymorphisms accurately and to a high resolution. Therefore, GWAS has become a powerful tool for detecting SNPs linked to traits of interests. To date, many studies have identified associations of DNA markers (microsatellite or SNPs) or genes of important economic traits in relevant aquaculture species by GWAS, including *C. gigas* (Chi et al., 2024; Du et al., 2023; Gutierrez et al., 2018; Wu et al., 2023; Yang et al., 2022). These studies show that growth, meat yield, and

disease resistance are generally controlled by a significant number of genes with minor effects.

This study established a linkage map and perform GWAS to determine SNPs markers linked to whole weight, shell shape and meat yield in *C. angulata*. Also, the linkage map will be used for identifying genome regions controlling traits of interest. High-density genotyping by sequencing technology was used to genotype all sampled oysters in this study.

2. Materials and method

2.1. Oyster collection

Crassostrea angulata samples were taken from selection program for improving high growth in Vietnam (Vu et al., 2020a, 2020b). Oyster mantle tissue samples were collected after approximately 9-months grow-out in the open ocean (Vu et al., 2021a, 2021b). Phenotypic measurements of the oysters were recorded, including whole weight (647 samples), shell shape (361 samples), and meat yield (490 samples). DNA samples were collected, originated from two generations of a selective breeding program and consisted of 188 oysters from 57 full-sib family of the first generation and 459 oysters of 33 full-sib family of the second generation. During the data collection for meat yield, individual oysters were also inspected for gonadal products by microscopy to identify their sex. All tissue samples were then stored in 80% ethanol and preserved at -80°C before transferring them for sequencing at Diversity Arrays Technology Pty. Ltd, Canberra, Australia.

2.2. Traits studied and phenotype measurements

Oyster shells were cleaned by pressure washing to remove algae and other fouling organisms before undertaking the phenotypic measurements. A digital balance with an accuracy of 0.01 was used to record whole weight. Oysters were cultured by strings hanging out on bamboo cages.

Cup ratio was used for characterization of shell shape in oysters. Dividing shell height by shell width is value of cup ratio (Walton et al., 2013). Shell dimensions were measured with callipers (accuracy of 0.01 mm).

Meat yield, measured as wet soft tissue weight, was determined with digital scales at an accuracy of 0.01g.

2.3. DNA extraction, library construction and sequencing

Tissue samples were extracted and purified following an established protocol at Diversity Array Technology Pty Ltd, Australia. Library construction and sequencing were detailed and described by Kilian et al. (2012) and Elshire et al. 2011).

2.4. Linkage analysis

We constructed linkage maps for the Portuguese oyster using the Lep-MAP3 software (Rastas, 2017) from the DArTseq genotype data of 647 individuals. First, the genotypes in a csv file were converted into Lep-MAP3 format using the genotypes2post.awk script. Then we validated the pedigree structure using Lep-MAP3's module IBD (identity-by-descent IBD > 0.9 for duplicates and about 0.5 for parent-offspring or between full-sibs) and constructed our final pedigree with 647 oysters. For 19 families, one or both parents were missing, and we added 29 "dummy" parents for those families to the pedigree.

This data consisted of two generations from a selective breeding of *C. angulata*. The linkage mapping followed the normal Lep-MAP3 pipeline (1) ParentCall2: Figures out parental genotypes, 2) Filtering2: Removes distorted markers, 3) SeparateChromosomes2: Puts markers into linkage groups, 4) JoinSingles2All: Adds additional markers to the linkage groups, 5) OrderMarkers2: Constructs the final linkage map). To cope with missing parents in the data, the parameters were chosen as suggested in the Lep-MAP3 documentation (ignoreParentOrder=1 in ParentCall2, minLod=0 in SeparateChromosomes2 and phasingIterations=3, refineParentOrder=1 in OrderMarkers2). Moreover, we used lod3Mode=1, lodLimit=15 in SeparateChromosomes2 to separate 10 linkage groups and lodLimit=12 (and 10 for second run), lodDifference=3 in JoinSingles2All to put a total of 5,129 markers into these 10 groups. The "calculateIntervals" parameter was used to store the marker position intervals (uncertainty in map position) for each linkage group.

2.5. Genome anchoring and linkage map improvement

We used bwa mem (Li, 2013) to map DArTseq sequence reads to the *C. angulata* genome (Qi et al., 2023), obtaining physical location for 4,085 (80% of 5,129) of the linkage map markers (Li, 2013). To improve the quality and the number of markers mapping to the genome, we also analysed the raw fastq files from DArTseq. The fastq files were mapped to the Portuguese oyster genome using bwa mem and the genotype calls were obtained using Lep-MAP3 pipeline (samtools (Danecek et al., 2021; Li, 2013) mpileup + Pileup2Likelihoods)

(Rastas, 2017). We called variants individually from each fastq and used IBD module to find 111 identical (>0.9) "individuals" (out of 574 fastqs) and joined these identical calls. Then we used the mapped genome coordinates to find the common markers in the csv file from DaRTseq and the new variants obtained from the fastq files. Taking the common markers, we could use IBD module to match individual names between the two datasets; with IBD>0.75 almost all individuals had a unique match (446 out of 463). All the matching individuals were offspring of the families (no parental data).

Then we produced a new linkage map using the same Lep-MAP3 pipeline as before. This map contained 19,475 markers (over 3x of the previous map) and all these had coordinates in the genome. Thus, we used only the new map to anchor the *C. angulata* genome (Qi et al., 2023) into pseudo-chromosomes. For this, we used Lep-Anchor (Rastas, 2020), running the Lep-Anchor wrapper (lepanchor_wrapper2.sh). For input data, we used the linkage map given as marker intervals, the raw nanopore reads (used in the assembly GCA_005518195.2, SRA:SRS4675889) aligned to the genome using minimap2 (Li et al., 2018) and contig-contig alignments calculated by the first two steps of HaploMerger2 (Huang et al., 2017), and another set of alignments computed as explained in the Lep-Anchor documentation using minimap2 (Li et al., 2018) and Lep-Anchor modules.

Based on the first run, read and contig alignments dominated the anchoring due to the low number (and sometimes quality) of linkage map markers. Therefore, we doubled each linkage map marker by expanding map intervals by 10 recombinations (if a marker had a map position between 100-110, a new marker was added with the same physical position but with linkage position between 90-120 = original interval expanded by 10 from both ends). The added markers and interval expansions increased map contribution to the anchoring. We further improved the linkage map using the anchored genome. We took all markers from each pseudo-chromosome and ran Lep-MAP3's SeparateChromosomes2 again within each chromosome using a lower lodLimit of 6. The final linkage map with 19,475 markers was obtained by evaluating the linkage map in the (anchored) physical order (parameter evaluateOrder in OrderMarkers2). These two improved maps were used for QTL mapping.

2.6. QTL mapping

We conducted QTL (Quantitative Trait Locus) mapping on two traits, the sex (male and female) and weight of the oysters. The (four-way) phased segregation data outputted from

Lep-MAP3 was associated with the traits using the generalised linear model "glm" function in the R language, like that in Li et al. (2018) (codes available when requested). We calculated the log odds (likelihood ratio) of the two glms with and without the segregation data separately for each family. The paternal and maternal segregation patterns were used as two additive covariates and combined to provide a dominance covariate (Li et al., 2018). Sex was used as a covariate in the glm for weight. The likelihood ratios for each family were multiplied together as the measure of association. Finally, we used a permutation test with 1000 replicates to obtain the significance of the multiplied log odds of the QTLs.

2.7. Genome-wide association study

Four different methods in GWAS used to identify SNPs including correlation test, principal component analysis, numeric regression, and mixed model. All data were analyzed by SVS suite software (Bozeman, 2010). Both single-SNP and multi-SNP analyses were used. The mixed model GWAS consisted of the fixed effects of generations (generations and sex) and a kinship matrix as a random effect. After quality control for GWAS analysis, the genotypic data from 13,048 markers was used to compute the kinship matrix with call rate of 50%. Basic statistics about the sequence data after quality control is shown in Supplementary 1. The description of mixed model for identifying SNP markers in this study is written in equation 1 below:

$$y = Xb + Ga + e \quad (1)$$

where *y* is the vector of observations for traits studied (whole weight, soft tissue weight, cup ratio), *X* is the incidence matrix regarding the fixed effects. *b* is the vector of all possible systematic fixed effects, including generation, sex. Vector *a* is the random animal additive genetic effect variable ~ (0, A σ_a^2) where *G* is the genomic relationship matrix calculated from the SNP markers, and e is the vector of residual effects ~ (0, $I\sigma_a^2$). Furthermore, we calculated the multiplicative inflation factor [λ = median (X_i^2)/0.456] where X_i^2 is the distribution of the Chi-square statistics over a set of markers or the actual association tests. The lambda value λ > 1 shows population substructure or genotyping errors (SNPs with λ > 1 were omitted). Multiple testing corrections used both Bonferroni correction and false discovery rate (FDR) methods to minimize type I errors (i.e., SNPs declared are significant, but the association is not present). In this study, there were about 13,048 SNPs analysed, the Bonferroni corrected significance level was set at 10^{-4}. However, the high probability of type II

errors is known to be linked to the Bonferroni test. The reliability of analysis is the false discovery rate (FDR) with a significant level of 0.05. When significant SNPs were detected, we obtained DNA sequences from DArT company and annotated these sequences in the NCBI to get characterized genes.

3. Results

3.1. Measurement data

The mean harvest whole weight of 647 samples was 51.1 g with a standard deviation of 16.1 g. The harvest whole weight of oysters of the same 12 months old ranged from 15.7 g to 69.9 g, showing a high level of variability. The cup ratio demonstrated a typical height/width/length ratio of 1:1.6:2.9. All phenotypic data such as whole weight, meat yield (soft tissue weight), cup ratio is presented in Table 1.

Table 1. Phenotypic measurements of whole weight, soft tissue weight, shell shape (cup ratio)

Traits	Unit	n	Mean	SD	Min	Max
Whole weight	g	647	51.07	16.13	15.71	69.90
Soft tissue weight	g	490	9.55	3.13	1.05	19.92
Cup ratio	Ratio	361	1.61	0.33	0.74	2.60

3.2. Linkage maps

Sex-averaged maps

A total of 19,475 SNP markers were mapped to 10 chromosome-level assemblies. The total average map length was 931.7 cM (Table 2). Number of SNPs per pseudo-chromosome in the linkage map ranged from 1,193 to 2,609 (mean = 1947.5) (Table 2). The chromosome lengths ranged from 70.3 to 133.2 cM (mean = 93.17). The average distance between markers ranged from 0.031 to 0.103 cM with an average of 0.0524. The sex-average marker map is presented in Figure 1.

Table 2. Summary statistics of the sex-averaged linkage map of *C. angulata*

Chromosome -level assemblies	Number of SNP markers	Map length (cM)	Average distance between markers (cM)	Anchored physical length (Mb)
1	2609	91.3	0.034	58.1

2	2511	91.9	0.037	57.7
3	2527	77.6	0.031	54.7
4	1985	87.8	0.044	52.1
5	2013	70.3	0.035	51.5
6	1981	90.0	0.045	55.3
7	1815	76.2	0.042	42.2
8	1554	101.0	0.065	46.6
9	1193	103.4	0.087	30.5
10	1287	133.2	0.103	53.3
Total	19475	931.7	0.524	586.0
Average	1947.5	93.17	0.0524	58.6

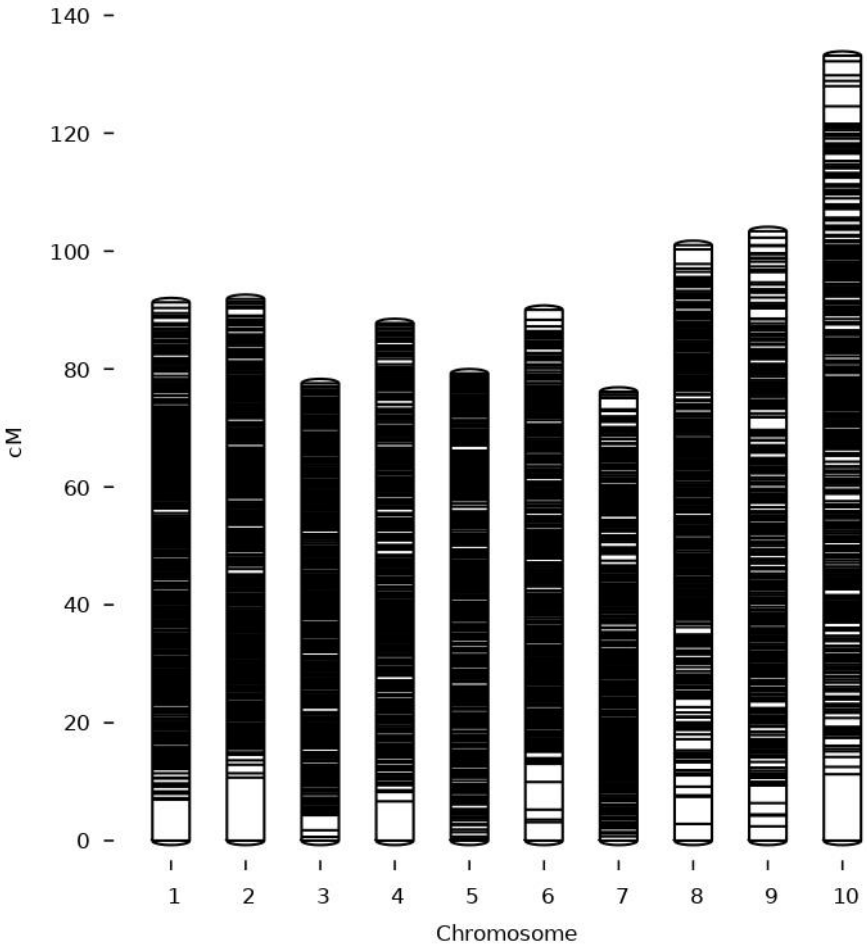


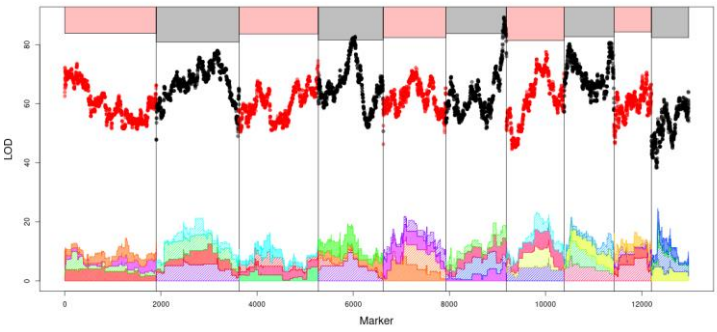
Figure 1. Sex averaged linkage map

Segregation distortion

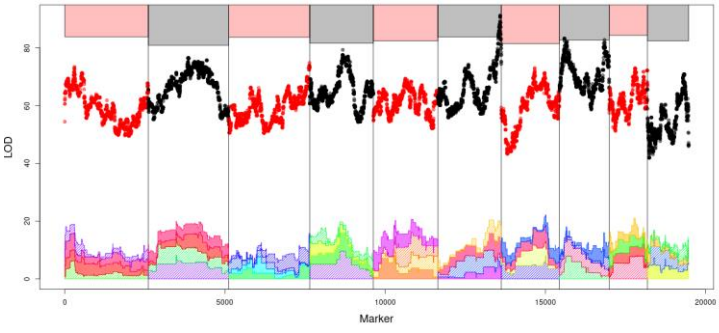
The segregation distortion, based on markers and the analysis, indicated considerable segregation (*P < 0.001*) in some families. The distorted loci were filtered before the linkage map was constructed. At this level, one or more families were filtered for 13,048 markers with at least one segregating family.

QTL mapping

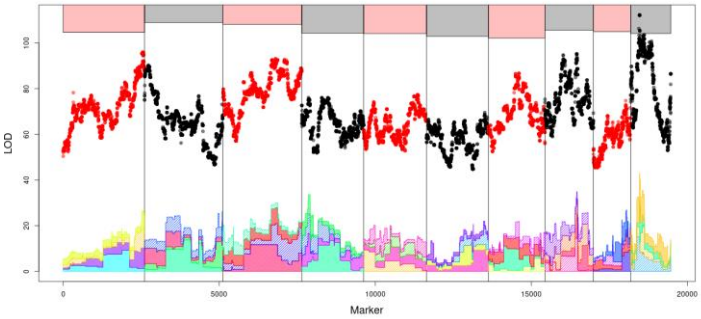
Significant (p < 0.05) QTL peaks found from the anchoring map and physical map were very similar. For the weight, there was only one peak in chromosome 10, and for sex, the highest peak was in chromosome 6. Additionally, the physical map had two extra sex-associated peaks in chromosome 8 and the anchoring map had an extra peak in chromosome 4 (Figure 2). We identified the location of each of these regions in the anchored genome as chromosome 4: 26042760-33763831, chromosome 6: 51702873-54673677, chromosome 8: 6521458-6554153 and 42797158-42871257 and chromosome 10: 17152944-18503092. The actual QTL regions could be slightly larger, especially in chromosome 8 where only a few markers spanned both peaks.



sex, anchoring map



sex, physical map



weight, anchoring map

Figure 2. QTL analysis as Manhattan plot for the sex (two upper panels) and weight (lower panel).

The upper and lower panels are showing the association for the anchoring map (independent of the physical marker order), the middle panel is for the map evaluated in the physical order. The LOD displayed for each marker is the multiply of single family likelihood ratios. The vertical lines separate the linkage groups as well as alternative colors of black and red for even and odd linkage groups. Significant peak levels obtained by permutation test are shown as the transparent (red or black) boxes above each linkage group. The four families with highest contribution to the LOD are shown below each group.

3.3. GWAS for growth, meat yield, and shell shape

Using a significance threshold of *P<0.0001* (after Bonferroni correction), six SNPs were significantly linked to harvest whole weight (Table 3), where two loci were mapped to chromosome 5, and the remaining ones were mapped to chromosome 6 and chromosome 10. Interestingly, these SNPs were also found in three other statistical methods (Table 4). The possible effects of the population structure and cryptic relatedness that may result in biases in the results were evaluated by Q-Q plot (Figure 3).

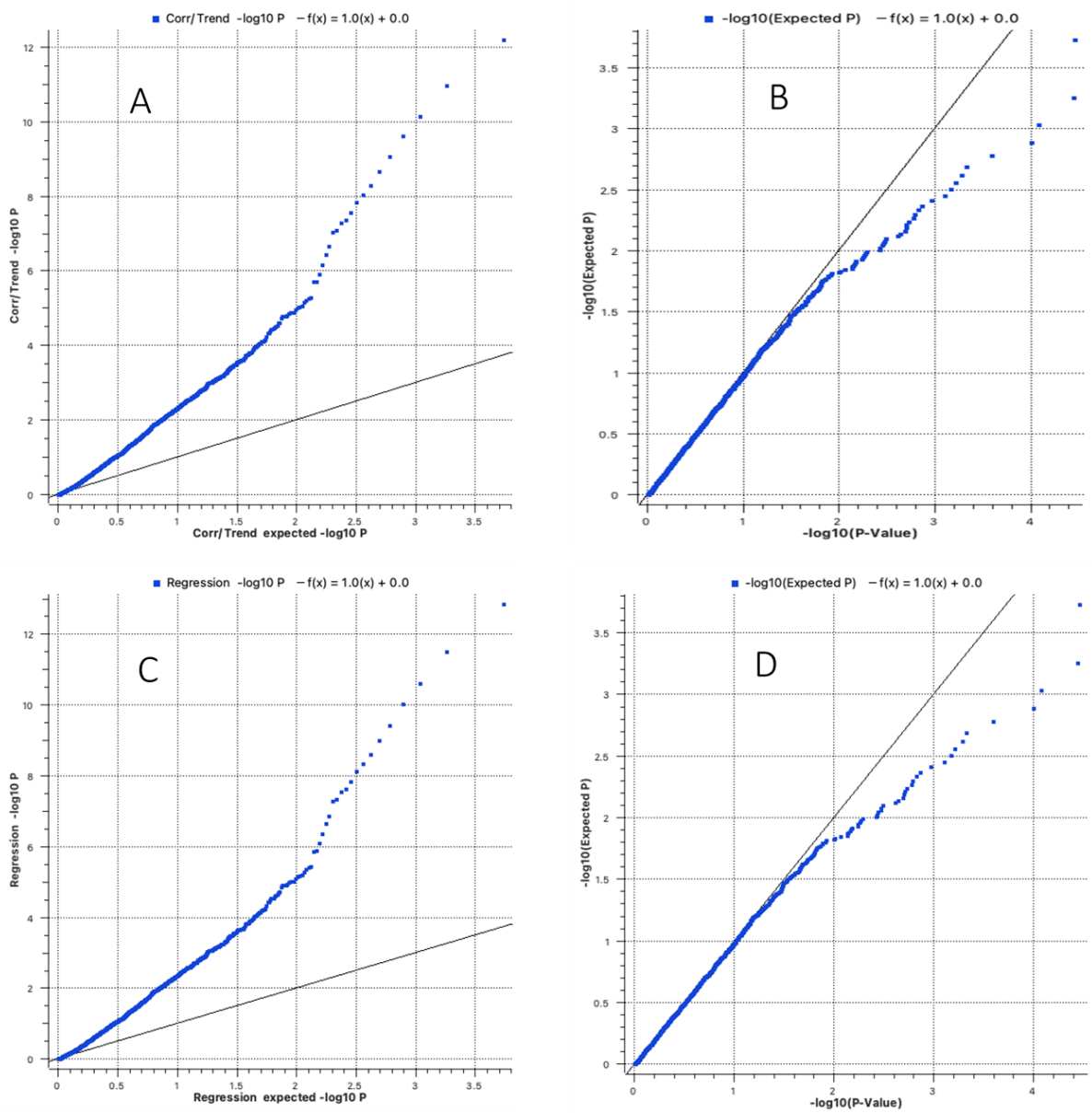


Figure 3. Quantile-quantile plot of P-values from single SNP genome wide association study for whole weight trait (***A***, correlation/trend test; ***B***, single locus mixed model; ***C***, numeric regression with corrections for PCAs, and ***D***: multiple locus mixed model)

Three significant SNPs were associated with meat yield (Figure 4, Table 3) and four SNPs were significantly associated with shell shape, explaining 9.2% and 8.2% of total genetic variance, respectively. Meanwhile, six SNPs/variants linked to whole weight explained 10.2% of the total genetic variance of whole weight trait. Interestingly, these SNPs were also detected from three other statistical methods (Table 4).

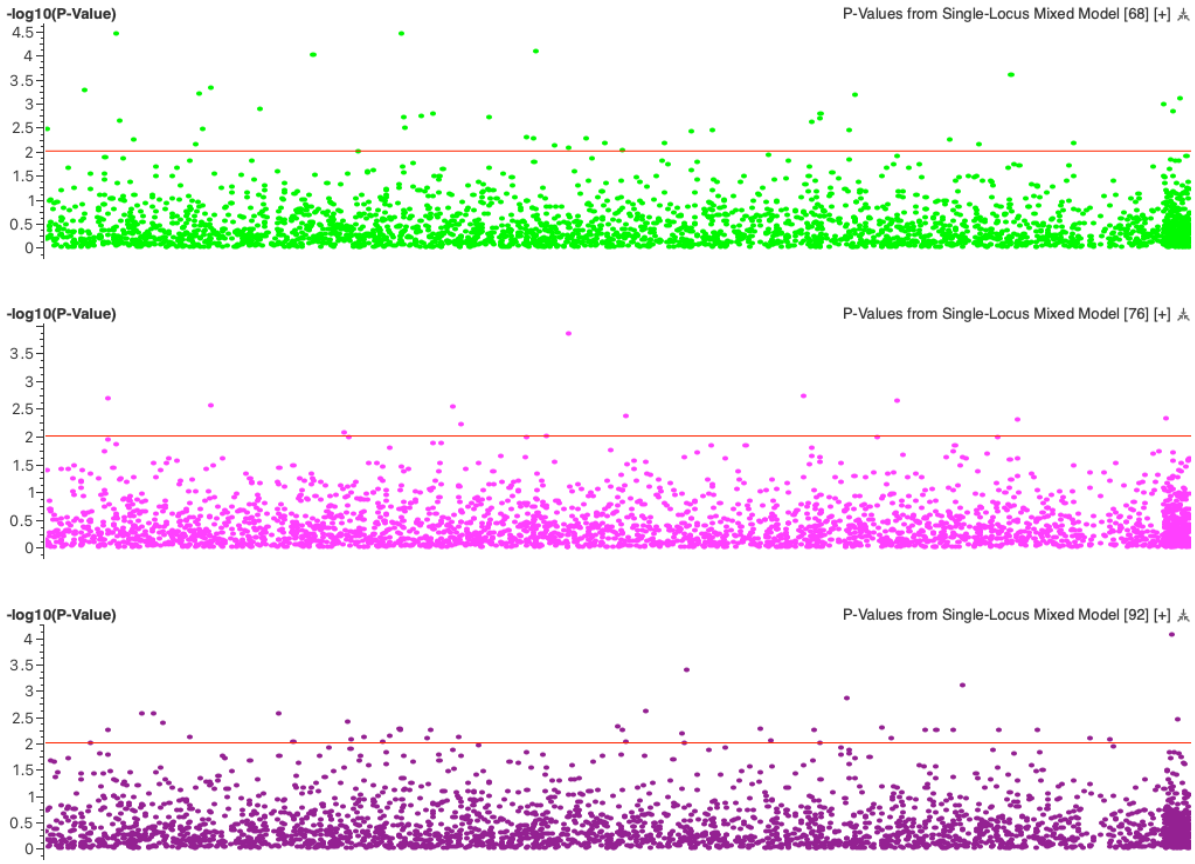


Figure 4. The Manhattan plot showing the -log$_{10}$(P-values) of SNPs on whole weight (top), wet soft tissue weight (middle), cup ratio (bottom) using the mixed model.

Table 3. Significant markers (*P <0.0001*) linked to whole weight, soft tissue weight and cup ratio using the mixed model

Trait	Marker	Position (bp)	P-value	Pseudo-chromosome	MAF
Whole weight	SNP#30057485	875	5.0E-06	5	0.013
	SNP#30061115	9958	3.7E-06	10	0.059
	SNP#30054712	4377	2.2E-05	5	0.114
	SNP#30048871	6036	5.3E-05	6	0.044
	SNP#30063907	13961	2.7E-04	6	0.036
	SNP#30062561	7614	8.4E-04	10	0.136

Meat yield (soft tissue weight)	SNP#30052842	6444	1.1E-04	1	0.034
	SNP#30063909	9333	2.5E-03	1	0.376
	SNP#30055957	10479	2.6E-03	1	0.025
Shell shape (cup ratio)	SNP#42707964	13861	1.8E-04	5	0.067
	SNP#30058304	7893	5.5E-04	2	0.028
	SNP#30048118	11286	9.3E-04	8	0.262
	SNP#30048208	10289	1.3E-04	2	0.037

MAF, minor allele frequency

Table 4. Number of SNPs (N) and their false discovery rates (FDRs) (%) for whole weight, soft tissue weight (meat yield) and cup ratio using the mixed model

Trait	Correlation		Regression with PCA corrections		Single locus mixed model		Multiple locus mixed model	
	N	*FDR*	*N*	*FDR*	*N*	*FDR*	*N*	*FDR*
Whole weight	28	0.9×10^{-4}	37	0.9×10^{-4}	11	0.79×10^{-4}	6	0.80×10^{-4}
Meat yield (soft tissue weight)	97	0.9×10^{-4}	106	0.9×10^{-4}	13	0.98×10^{-4}	12	0.87×10^{-4}
Shell shape (cup ratio)	16	2.3×10^{-4}	57	0.9×10^{-4}	12	0.97×10^{-4}	10	0.85×10^{-4}

Genes with known functions linked to identified SNPs

Along with the determination of significant SNPs linked to whole weight, meat yield, and shell shape, our study found that these significant SNP markers can be used to determine positions of genes with known growth and immune functions, as well as genes encoding for proteins involved in growth function. Genes that are well characterized and potentially regulated growth including neuroligin-4 (Y-linked-like), transmembrane protein 45B. Therefore, the genes from significant SNP markers seem to linked to genes which regulate directly whole weight, soft tissue weight, and shell shape (cup ratio).

4. Discussion

This study successfully established a high-density linkage map for the Portuguese oyster, *C. angulata* which can serve as a reference for future genomic research in *C. angulata*. GWAS analyses detected SNPs linked to a commercially important traits that can be used for the further development of SNPs panel applications when more samples are collected. Candidate genes identified linked to commercial traits can be useful for commercial implementation marker-assisted selection in the *C. angulata* industry.

Linkage map

The quality of the constructed linkage map indicates that the used DarTSeq data is of good quality and evenly distributed across the genome. The map enabled us to anchor 86% of the genome into 10 pseudo-chromosomes. This first dense SNP-based linkage map for *C. angulata*, consisting of 19,475 markers mapped to 10 chromosomes with 931.6 cM span (average marker interval of 0.0524 cM). The average marker interval in this study was more denser than that of genetic linkage map (an average of 0.8 cM) in a hybrid family of oysters (*C. gigas* x *C. angulata*) (Wang et al., 2016). This can be caused by DNA sequencing density or platform carried out for these two researches. The developed SNPs panel from this study can be applied to other populations, and the collection of samples from different origins is essential in qualifying its utility. In constructing the linkage map, we removed the distorted markers with $p<0.001$ to prevent any possible bias that may originate from genotyping errors. Segregation distortion is reported due to the accumulation of recessive deleterious mutations, genetic load, duplicated genes, transposable elements, and unusual meiotic segregation distortion (Liao et al., 2024). Gametic incompatibility or reduced hybrid viability also causes uneven transmission of alternate alleles, which is frequently caused by disrupted genetic interactions among loci of parental lineages, leading to the non-random elimination of particular allelic combinations (Liu et al., 2002). The low segregation distortion rate in this study may have been due to the absence of the aforementioned factors in this study. Furthermore, the uneven distribution of the distorted markers among chromosomes also suggests that marker distortion was not caused by technical limitations or other typing errors. The segregation distortion rate was not reported in recent studies for mollusc species that used similar genotyping by sequencing technology (Hollenbeck and Johnston, 2018). Nevertheless, segregation distortion rate in *C. gigas* has been reported as moderate to high,

ranging from 17 to 66% (Wang et al., 2016). In the study of (Jourdan et al., 2023), large differentiation in the distribution of SNPs across the genome were observed in the marker density maps. This could result in missing QTLs in some regions but this might be overcome either by using an existing high-density array like the 190K array (Qi et al., 2017; Wang et al., 2016) or by developing a new, more optimised array.

Genome-Wide Association Study (GWAS)

Along with the determination of significant SNPs linked to whole weight, meat yield, and shell shape, our study indicated that these significant SNPs markers are well located to determine the position of genes with known growth and immune functions, as well as in genes encoding for proteins involved in growth. These SNPs explained 10.2% of the total genetic variance of whole weight and 9.2% of the total genetic variance for meat yield. The discoveries from this study were not affected by the population stratification because there were no correlations between the population structure with the phenotypes (Vu et al., 2021). Genes that are well characterized and potentially related to whole weight include neuroligin-4 (Y-linked-like), and transmembrane protein 45B. This suggests that whole weight in *C. angulata* involves complex processes, possibly controlled by many genes with relatively smaller effects. Their identification would require much larger samples size and denser markers (Spencer et al., 2009). Understanding the genetic architecture of quantitative complex traits in conjunction with conventional selective breeding method can improve selection response for complex traits such as meat yields and shell shape. In addition, the phenotypic correlations between whole weight and soft tissue weight (meat yield) were high and favourable (Vu et al., 2019), possibly controlled by many genes with small effects. Similarly, genetic correlation was high and favourable between whole weight and soft tissue weight but there was no significant genetic correlation between whole weight and cup ratio (Vu et al., 2021). Among the studied traits, shell shape were also mentioned with other indexes such as height/ width/ length ratio (Evans and Langdon, 2003). Whole genome resequencing is expected to have greater power to detect/identify SNPs markers that significantly contribute to variation in quantitative complex traits as included in our analysis here and SNP chip development will be more useful for *C. angulata* industry (Peñaloza et al., 2022; Zhang et al., 2024). Genes found from this

study was also agreement with those discovered in Wang et al. (2016) and were responsible for metabolism.

Conclusion

A high-density linkage map was built for Portuguese oyster, *Crassostrea angulata*. A total of 19,475 markers were mapped to 10 chromosomes. Six SNPs were significantly linked to the whole weight. These SNPs consisted of about 10.2% of the total genetic variance of whole weight. Further study should be carried out to discover their biological functions regulating growth traits in Portuguese oyster, *C. angulata*.

Funding

This study was funded by the Australian Centre for International Agricultural Research, Australia via the project: Enhancing bivalve production in northern Vietnam and Australia (FIS/2010/100). PR is funded by the Academy of Finland, grant no. 343656.

Reference

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

Griot, R., Allal, F., Phocas, F., Brard-Fudulea, S., Morvezen, R., Bestin, A., Haffray, P., François, Y., Morin, T., and Poncet, C. (2021). Genome-wide association studies for resistance to viral nervous necrosis in three populations of European sea bass

(*Dicentrarchus labrax*) using a novel 57k SNP array DlabChip. *Aquaculture* **530**, 735930.

Gutierrez, A. P., Bean, T. P., Hooper, C., Stenton, C. A., Sanders, M. B., Paley, R. K., Rastas, P., Bryrom, M., Matika, O., and Houston, R. D. (2018). A genome-wide association study for host resistance to ostreid herpesvirus in pacific oysters (*Crassostrea gigas*). *G3: Genes, Genomes, Genetics* **8**, 1273-1280.

Huvet, A., Fabioux, C., McCombie, H., Lapègue, S., Boudry, P. (2004). Natural hybridization between genetically differentiated populations of *Crassostrea gigas* and *C. angulata* highlighted by sequence variation in flanking regions of a microsatellite locus. Mar. Ecol. Prog. Ser. **272**, 141–152

Jiao, W., Fu, X., Li, J., Li, L., Feng, L., Lv, J., Zhang, L., Wang, X., Li, Y., and Hou, R. (2014). Large-scale development of gene-associated single-nucleotide polymorphism markers for molluscan population genomic, comparative genomic, and genome-wide association studies. *DNA Research* **21**, 183-193.

Jin, Y., Zhou, T., Geng, X., Liu, S., Chen, A., Yao, J., Jiang, C., Tan, S., Su, B., and Liu, Z. (2017). A genome-wide association study of heat stress-associated SNP s in catfish. *Animal genetics* **48**, 233-236.

Kilian, A., Wenzl, P., Huttner, E., Carling, J., Xia, L., Blois, H., Caig, V., Heller-Uszynska, K., Jaccoud, D., and Hopper, C. (2012). Diversity arrays technology: a generic genome profiling technology on open platforms. *In* "Data production and analysis in population genomics", pp. 67-89. Springer.

Qi, H., Cong, R., Wang, Y., LI, L. and Zhang, G. (2023). Construction and analysis of the chromosome-level haplotype-resolved genomes of two Crassostrea oyster congeners: *Crassostrea angulata* and *Crassostrea gigas. GigaScience **12**, 1–14.*

Lapègue, S., Heurtebise, S., Cornette, F., Guichoux, E., and Gagnaire, P.-A. (2020). Genetic characterization of cupped oyster resources in Europe using informative Single Nucleotide Polymorphism (SNP) panels. *Genes* **11**, 451.

Lawrence, D., and Scott, G. (1982). The determination and use of condition index of oysters. *Estuaries* **5**, 23-27.

Li, N., Zhou, T., Geng, X., Jin, Y., Wang, X., Liu, S., Xu, X., Gao, D., Li, Q., and Liu, Z. (2018). Identification of novel genes significantly affecting growth in catfish through GWAS analysis. *Molecular Genetics Genomics* **293**, 587-599.

Meng, J., Song, K., Li, C., Liu, S., Shi, R., Li, B., Wang, T., Li, A., Que, H., and Li, L. (2019). Genome-wide association analysis of nutrient traits in the oyster *Crassostrea gigas*: genetic effect and interaction network. *BMC genomics* **20**, 1-14.

Meng, J., Wang, W., Shi, R., Song, K., Li, L., Que, H., and Zhang, G. (2020). Identification of SNPs involved in Zn and Cu accumulation in the Pacific oyster (*Crassostrea gigas*) by genome-wide association analysis. *Ecotoxicology Environmental Safety* **192**, 110208.

O'Connor, W., Dove, M., O'Connor, S., In, V. V., Lien, V. T. N., and Van, P. T. (2019). Enhancing bivalve production in northern Vietnam and Australia. *Final Report. Australian Centre for International Agricultural Research. FIS/2010/100*.

Tsai, H.-Y., Hamilton, A., Tinch, A. E., Guy, D. R., Gharbi, K., Stear, M. J., Matika, O., Bishop, S. C., and Houston, R. D. (2015). Genome wide association and genomic prediction for growth traits in juvenile farmed Atlantic salmon using a high-density SNP array. *BMC Genomics* **16**, 1-9.

Vu, S. V., Gondro, C., Nguyen, N. T., Gilmour, A. R., Tearle, R., Knibb, W., Dove, M., Vu, I. V., Khuong, L. D., and O'Connor, W. (2021a). Prediction Accuracies of Genomic Selection

for Nine Commercially Important Traits in the Portuguese Oyster (*Crassostrea angulata*) Using DArT-Seq Technology. *Genes* **12**, 210.

Vu, S. V., Premachandra, H., O'Connor, W., Nguyen, N. T., Dove, M., Van Vu, I., Le, T. S., Vendrami, D. L., and Knibb, W. (2021b). Development of SNP parentage assignment in the Portuguese oyster *Crassostrea angulata*. *Aquaculture Reports* **19**, 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.

Vu, V. I., Vu, V. S., O'Connor, W., Phan, T. V., Dove, M., Knibb, W., and Nguyen, N. H. (2017b). Are strain genetic effect and heterosis expression altered with culture system and rearing environment in the Portuguese oyster (*Crassostrea angulata*)? *Aquaculture Research* **48**, 4058-4069.

Vu, S. V., Knibb, W., Gondro, C., Subramanian, S., Nguyen, N. T., Alam, M. & O'Connor, W. (2021). Genomic prediction for whole weight, body shape, meat yield, and color traits in the Portuguese oyster *Crassostrea angulata*. *Frontiers in Genetics*, *12*, 661276.

Vu, V. S., Knibb, W., Nguyen, T. H. N., Vu, V. I., O'Connor, W., Dove, M., and Nguyen, N. H. (2019). First breeding program of the Portuguese oyster *Crassostrea angulata* demonstrated significant selection response in traits of economic importance. *Aquaculture* **518**, 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.

Wang, J., Qi, L., Zhang, J., Kong, L., and Yu, Y. (2021). High macro-collinearity between *Crassostrea angulata* and *C. gigas* genomes was revealed by comparative genetic mapping with transferable EST-SNP markers. *Aquaculture* **545**, 737183.

Wang, J., Li, L. & Zhang, G. (2016). A high-density SNP genetic linkage map and QTL analysis of growth-related traits in a hybrid family of oysters (*Crassostrea gigas*× *Crassostrea angulata*) using genotyping-by-sequencing. *G3: Genes, Genomes, Genetics*, *6*(5), 1417-1426.

Yang, Y., Wu, L., Wu, X., Li, B., Huang, W., Weng, Z., Lin, Z., Song, L., Guo, Y., and Meng, Z. (2020). Identification of candidate growth-related SNPs and genes using GWAS in brown-marbled grouper (*Epinephelus fuscoguttatus*). *Marine Biotechnology* **22**, 153-166.

Yu, Y., Wang, Q., Zhang, Q., Luo, Z., Wang, Y., Zhang, X., Huang, H., Xiang, J., and Li, F. (2019). Genome scan for genomic regions and genes associated with growth trait in pacific white shrimp *Litopeneaus vannamei*. *Marine Biotechnology* **21**, 374-383.

Zeng, Z., Tan, Q., Huang, Z., Shi, B., and Ke, C. (2019). Differential Gene expression related to morphological variation in the adductor muscle tissues of diploid and triploid fujian oysters, *Crassostrea angulata*. *Aquaculture Research* **50**, 3567-3578.

Zhou, Z., Chen, L., Dong, C., Peng, W., Kong, S., Sun, J., Pu, F., Chen, B., Feng, J., and Xu, P. (2018). Genome-scale association study of abnormal scale pattern in yellow river carp identified previously known causative gene in European mirror carp. *Marine Biotechnology* **20**, 573-583.

Supplementary

Supplementary 1. Basic statistics about the sequence data.

SNP statistics	Mean	Range
Total SNPs	19,475	
Average Count Snp	23.7	2.5 – 295.4
Average Count Ref	33.6	2.5 – 451.5
Frequency Hom Snp (% samples which score homozygote)	0.22	0 – 1
Frequency Hets (% samples which score heterozygote)	0.11	0 – 0.68
PIC SNP	0.28	0 – 0.5