VALUE-ADDING TO PEARL OYSTERS

This Food Agility CRC project used digital technology to improve the genetic selection of Akoya oysters.

By Dean Jerry, Nga Vu, Carla Ewels, and Mostafa Rahimi Azghadi

foodagility

JAMES COOK UNIVERSITY AUSTRALIA

BROKEN BAY PEARL FARM

Food Agility Cooperative Research Centre Project No. FA140

ISBN 978-1-7646726-4-1

Project Name: Digital phenotyping to improve the efficiency of genomic selection in Akoya oysters

Project No: FA140

Citation: Dean Jerry, Nga Vu, Carla Ewels, and Mostafa Rahimi Azghadi (2026), Digital phenotyping to improve the efficiency of genomic selection in Akoya oysters, Food Agility CRC, Sydney.

IMPORTANT INFORMATION

Author contact details
Name: Dean Jerry
Phone: +61 747815586
Email: dean.jerry@jcu.edu.au

Food Agility CRC contact details
L4, 15 Martin Place
SYDNEY NSW, 2000

+61 2 8001 6119
hello@foodagility.com
www.foodagility.com

Electronically published by Food Agility CRC at www.foodagility.com in August 2026.

The Food Agility CRC is funded under the Australian Government Cooperative Research Centres Program.

Australian Government
Department of Industry, Science and Resources
AusIndustry
Cooperative Research Centres Program

ACKNOWLEDGEMENTS

The project team gratefully acknowledges the financial support provided by Food Agility CRC, Cygnet Bay Pearls Unit Trust (Broken Bay Pearls), and James Cook University (JCU), whose partnership made this project possible. We also sincerely thank these organisations for their ongoing support, collaboration, and valuable feedback throughout the project.

We extend our sincere appreciation to the students and staff of the Aquaculture Genetics Team, as well as Engineering and Data Science at JCU, for their assistance with sample collection, laboratory work, and data generation and processing. We also thank the staff of Broken Bay Pearl Farm, Harvest Road, and the East 33 Hatchery for providing oyster samples and supporting field collections. We gratefully acknowledge Mick Schaefer, David Lamb, Chris Komorek, and Lucy Hickman for their project management, guidance, and ongoing support throughout the project.

Table of Contents

PROJECT DESCRIPTION

Akoya pearl oysters are an emerging high-value aquaculture species with strong potential for selective breeding to improve meat yield and product quality. To support the development of advanced breeding programs, this project developed industry-scalable artificial intelligence (AI)-based computer vision, sensor-driven phenotyping technologies, genomic tools, databases, and genomic prediction methodologies. These technologies enabled the rapid and accurate collection of phenotypic and genomic data, which were integrated into genomic prediction models to estimate breeding values for commercially important traits. Collectively, the project provides the technological platform needed to implement industrial-scale selective breeding programs, accelerating genetic improvement and enhancing the productivity, product quality, and profitability of the Australian Akoya oyster industry.

Lean Canvas

Project Name: FA140 - Digital phenotyping to improve the efficiency of genomic selection in Akoya oysters

1. End-Users	***Primary end-users*** • Australian Akoya oyster producers and selective breeding programs • Broken Bay Pearls • Harvest Road ***Early adopters*** • Commercial Akoya oyster producers • Emerging Australian Akoya farming enterprises • Future selective breeding programs for Akoya oysters ***Other value chain beneficiaries*** • Hatcheries and seed producers implementing improved broodstock selection • Genotyping service providers delivering SNP-based breeding services • AI and precision aquaculture technology providers • NIR technology developers for seafood quality assessment • Seafood processors, retailers, and restaurants through improved product quality and consistency • Consumers through improved product quality and sustainable production

2. End-User Problem(s)	• Akoya oyster is a high-value emerging aquaculture species with limited genetic improvement. • Industry requires efficient selective breeding to improve meat yield, product quality, and environmental resilience. • Large-scale breeding requires accurate, high-throughput collection of phenotype and genotype data. • Conventional phenotyping is labour-intensive, time-consuming, and limits the number of oysters that can be evaluated. **Today's alternative solutions** • Manual collection of data for oyster phenotypes which is slow, labour-intensive and often cost prohibitive at industrial-scales. • Relying on classical genetic selection approaches rather than advanced genomic prediction and machine learning. • Can be destructive to get some commercially important phenotypes limiting use as future broodstock.
3. Unique Value Proposition	• Rapid, objective, and high-throughput phenotyping for selective breeding. • Cost-effective genomic tools for routine breeding applications. • Improved selection accuracy for growth and meat quality traits. • Foundation for industrial-scale selective breeding of Australian Akoya oysters.
4. Our Solution	• Developed AI-based digital phenotyping for automated measurement of growth and meat traits. • Developed NIR models for rapid assessment of oyster meat quality. • Developed and validated a custom targeted SNP genotyping array. • Developed genomic prediction selection models for growth and meat quality traits.
5. Key Results	• AI-based phenotyping accurately measured shell dimensions (R^2 > 0.94) and predicted whole oyster weight (R^2 = 0.79) and meat weight (R^2 = 0.84). • Freeze-dried NIR models achieved acceptable prediction of glycogen content (R^2 = 0.74), demonstrating potential for rapid meat quality assessment. • A custom targeted SNP array achieved >95% genotyping success and enabled parentage assignment, population genetic analyses, and genomic selection. • Genomic prediction models generated genomic breeding values for growth and meat quality traits, supporting improved selection decisions. • Established the first integrated phenotype, genotype, and genomic resource for Australian Akoya oyster breeding.

6. Partners + Team	• JCU – Prof Dean Jerry, Prof Mostafa Rahimiazghadi, Dr Carla Ewells, Dr Nga Vu, Dr Cecile Massault, Julie Goldsbury, Anna Wilson, Vinu Sankar, Lenel Ahwan. • Broken Bay Pearls - James Brown, Duncan Smith, Andrew Williamson.
7. Pathways to Scale / Impact	• The project outcomes will allow companies to scale up the number of phenotypic and genotype records they can collect within a breeding program, increasing ability to predict genetic merit of broodstock more accurately. • It also allows rapid grading
8. Data Sets	• Akoya oyster SNP discovery database (WGS and ddRAD) • Broodstock and offspring genotype database (target SNP array genotypes) • Phenotype database (growth, meat quality, and AI-derived traits) • AI image library (annotated oyster images for machine learning) • Genomic prediction database (genomic estimated breeding values for growth and meat quality traits) • Population genetics and parentage database (population diversity and structure, relatedness, and pedigree assignments).
9. Research Questions	• Can machine learning (ML) be trained to accurately predict phenotypes of commercial interest? • Will NIRs provide fast and accurate real-time information on nutritional quality of oysters? • Does ML genetic merit prediction improve accuracy over conventional approaches?
10. Highest Risk Assumptions	• NIRs can accurately predict glycogen, lipid and protein composition • Machine learning can measure commercial phenotypes based on digital image capture with low error compared to manual measurements. • Machine learning increases the prediction accuracy of estimated breeding values of Akoya Oysters over conventional statistical approaches.

PROJECT PARTNERS

This project is a partnership between:

foodagility	Food Agility (Innovation Partner) brokers, designs, and delivers innovation programs for the Australian agrifood industry. As a project partner and funder, Food Agility supported the co-design, development, implementation, management, and administration of the project.
BROKEN BAY PEARL FARM	Broken Bay Pearls Pty Ltd (BPP) is an aquaculture company operating in the Hawkesbury River, New South Wales, producing both pearls and edible Akoya oysters. As the industry partner, BPP provided access to commercial farming operations, hatchery facilities, oyster families, and production data to support phenotyping, genotyping, and the development and validation of genomic tools. BPP also provided industry input to help ensure the project's outcomes were relevant and applicable to commercial production.
JAMES COOK UNIVERSITY AUSTRALIA	James Cook University (JCU) is Australia's leading university for tropical aquaculture research and training. As the research partner, JCU led the design, implementation, and overall management of the project. JCU contributed expertise in aquaculture genomics, quantitative genetics, and artificial intelligence, and undertook the development of the SNP array, AI-assisted computer vision and sensor phenotyping technologies, and genomic prediction models, supported by its state-of-the-art molecular and engineering research facilities.

EXECUTIVE SUMMARY

Overview

The Australian Akoya oyster industry is an emerging aquaculture sector with strong potential for premium seafood production. However, selective breeding has been constrained by the lack of efficient technologies for large-scale phenotyping and cost-effective genomic analysis. This project established an integrated breeding platform combining artificial intelligence (AI), near-infrared spectroscopy (NIR), genomic technologies, and genomic prediction to accelerate genetic improvement of Australian Akoya oysters.

Major project outcomes included computer-vision AI-based phenotyping tools for rapid measurement of growth and oyster meat traits, NIR prediction models for oyster meat quality, a custom targeted single nucleotide polymorphism (SNP) genotyping array, and genomic prediction selection models for growth and meat quality traits. Together, these technologies establish the genomic resources and decision-support tools required for industrial-scale selective breeding.

Why is the research important?

This research addresses one of the major barriers to developing an effective selective breeding program for Australian Akoya oysters: the ability to collect accurate phenotype and genotype data at commercial scales. By reducing reliance on labour-intensive manual measurements and providing cost-effective genomic tools, the project enables faster, more objective, and data-driven selection of broodstock for commercially important traits.

The project also generated the first integrated dataset linking growth and meat-quality measurements, SNP genotypes, and genomic prediction for Australian Akoya oysters, providing a foundation for long-term genetic improvement of growth, meat quality, and production efficiency.

Who will benefit from this research and how?

The primary beneficiaries are Australian Akoya oyster producers and hatcheries, who will benefit from improved breeding efficiency, more accurate selection decisions, and faster genetic gain for commercially important traits. Genotyping service providers and aquaculture technology developers can utilise the SNP array and AI-based phenotyping tools to support commercial breeding programs. Researchers and future breeding programs will benefit from the comprehensive genomic and phenotypic datasets generated by the project, while

consumers will ultimately benefit from improved product quality, consistency, and sustainability of Australian Akoya oyster production.

Key Results

- AI-enabled computer-vision phenotyping accurately measured oyster shell dimensions ($R^2 > 0.94$) and predicted whole oyster weight ($R^2 = 0.79$) and meat yield ($R^2 = 0.84$), enabling rapid, objective, and high-throughput phenotyping for selective breeding based on digital images.
- NIR spectroscopy demonstrated potential for rapid prediction of glycogen in freeze-dried samples, while further development is required for fresh tissue and lipid prediction.
- A custom targeted SNP genotyping array with >95% genotyping success was developed and validated, enabling routine parentage assignment, population genetic analyses, and power to predict genomic breeding values for the practice of genomic selection.
- The first comprehensive phenotype, genotype, understanding of the genetic basis of commercial traits of importance, and genomic prediction methodologies to support Australian Akoya oyster breeding were developed.
- Integrated AI and genomic technologies provide the foundation for industrial-scale selective breeding of Australian Akoya oysters.

INDUSTRY IMPACT METRICS

The key industry impacts include:

	**Impact metric 1**	Established AI-based high-throughput phenotyping capability for automated measurement of growth traits from digital images.
	**Impact metric 2**	Demonstrated the potential of NIR spectroscopy for rapid assessment of oyster meat quality traits, providing a foundation for future industry applications following further
	**Impact metric 3**	Established a validated targeted SNP genotyping platform to support parentage assignment, population genetics, and genomic selection
	**Impact metric 4**	Established the genomic resources and genomic selection capability required to support Australia's first selective breeding program for Akoya oysters

IMPACT STATEMENTS

INDUSTRY PARTNER

Broken Bay Pearl

James Brown and Duncan Smith, Broken Bay Pearls

For Broken Bay Pearls, this project has converted a long-term breeding ambition into a practical pathway for commercial application. By providing access to our farm, hatchery, oyster families and production data, we helped ensure that the technologies were developed around real operating conditions and the traits that matter to producers. The resulting AI-based image tools can measure shell dimensions with high accuracy and predict whole-oyster and meat weight, creating an opportunity to assess far more animals than is feasible through manual measurement alone. This could reduce labour, improve consistency and allow valuable broodstock to be evaluated using objective data. The custom SNP panel and genomic prediction models are equally important. They give industry a reliable platform for assigning parentage, managing relatedness and estimating breeding values for growth and meat-quality traits, even where traditional pedigree records are incomplete. Together, these capabilities will support better broodstock decisions, faster genetic improvement and a more structured approach to scaling Akoya oyster production. The project has also clarified where further work is needed. NIR spectroscopy showed promise for predicting glycogen in freeze-dried tissue, but it is not yet suitable for routine fresh-tissue assessment on farm. Our priority is therefore to continue validating the most mature tools under commercial conditions, expand the reference population and integrate routine imaging, genotyping and genomic evaluation into future breeding cycles. The partnership with James Cook University and Food Agility has built a strong technical foundation and strengthened industry capability. With continued collaboration and investment, these outcomes can improve productivity, product consistency and the long-term competitiveness of Australia's emerging Akoya oyster sector.

RESEARCH PARTNER

Prof Dean Jerry, Director ARC Research Hub for Supercharging Tropical Aquaculture through Genetic Solutions, James Cook University

James Cook University is proud to have led the development of innovative genomic and digital technologies to support selective breeding in the Australian Akoya oyster industry. Working closely with industry partners, this project developed the key tools required for modern breeding, including AI-based phenotyping, a custom SNP genotyping panel, and genomic prediction models for economically important traits.

These outcomes provide the foundation for implementing genomic-assisted selective breeding in Australian Akoya oysters, enabling more accurate breeding decisions and supporting long-term improvements in productivity and product quality. The project also established valuable genomic resources and integrated datasets that will support future breeding and research activities.

This project highlights the value of strong collaboration between research and industry to address emerging challenges in aquaculture. Continued investment and partnership will help translate these technologies into commercial breeding programs and strengthen the sustainability and competitiveness of the Australian Akoya oyster industry.

INNOVATION PARTNER

Food Agility CRC
David Lamb, Chief Scientist

Food Agility is proud to have supported this project. Breeding and utilizing Akoya pearl oysters specifically for eating is a novel, and emerging aquaculture sector in Australia. Typically farmed for luxury pearls, utilizing the oyster as a dedicated commercial food product is a diversification for the Australian seafood market. The development of genetics-based 'breeding values', along the lines of estimated breeding values ('EBVs) used for our well-established livestock industries, often comes later in the evolution of a food sector. It is exciting to be part of such foundational work undertaken right at the initial stages of an industry's establishment and growth.

END-USER PROFILE

Interview with James Brown and Duncan Smith, Broken Bay Pearls

Q: What problem were you trying to solve?

Broken Bay Pearls: "Akoya oysters are a high-value crop, but our ability to improve each generation has been limited by the amount of reliable information we can collect. Measuring shell size and weight by hand is slow and labour-intensive, particularly when thousands of oysters need to be assessed. Some important traits, such as meat yield and quality, have traditionally required oysters to be opened, which means potentially valuable broodstock cannot be retained. Farm production also mixes families over time, so conventional records alone do not always tell us which oysters are related or which broodstock produced the best offspring."

Q: Why is this project valuable to the farm?

Broken Bay Pearls: "The project gives us a practical pathway from visual assessment to evidence-based breeding. The AI image tools can measure shell dimensions accurately and predict whole-oyster and meat weight, allowing many more animals to be assessed consistently and with less manual handling. The SNP panel helps us assign parentage, understand relatedness and reconstruct families, while genomic breeding values allow us to compare animals on their genetic potential rather than appearance alone. Together, these tools can support better broodstock selection, reduce the risk of inbreeding and accelerate improvement in growth, meat yield and product quality. For the business, the long-term value is a more predictable crop, more efficient use of labour and stronger confidence when investing in future breeding cycles. The NIR work also showed where more development is needed: glycogen prediction was promising in freeze-dried tissue, but fresh-tissue assessment is not yet ready for routine farm use. That clarity helps us focus investment on the tools that are closest to commercial adoption while continuing to refine the others."



OBJECTIVES

The objectives of the FA140 project were as follows:

Objective 1: Develop a multi-view computer vision pipeline using deep learning-based image segmentation to automatically measure Akoya oyster phenotypic traits from digital images. This technology enables the rapid acquisition of phenotype data at industry scale to support selective breeding programs and potential future product grading applications.

Objective 2: Develop near-field infrared (NIR) prediction models to estimate oyster meat quality traits, including glycogen and lipid content. These models enable the rapid, non-invasive assessment of chemical-based quality traits at industry scale, supporting selective breeding programs and potential future product grading applications.

Objective 3: Develop and evaluate a custom Allegro SNP genotyping panel to support genetic improvement in the Australian Akoya oyster. This involved selecting informative SNPs using aquaculture Akoya oyster samples from New South Wales (NSW) and Western Australia (WA) and validating their suitability for selective breeding applications and genomic selection.

Objective 4: Optimize genomic prediction models for predicting breeding values of growth and meat quality traits. The outcomes are used to identify the most accurate and efficient genomic prediction framework for potential implementation in Akoya oyster breeding programs.

METHODOLOGY

Objective 1: Develop a multi-view computer vision pipeline using deep learning-based image segmentation to automatically measure Akoya oyster phenotypic traits from digital images

The initial dataset comprised 803 pearl oysters collected from the Broken Bay Pearl Farm during three collection periods: November 2024 (n = 152), June 2025 (n = 140), and December 2025 (n = 511). Ground-truthed (GT) phenotypic measurements, including whole oyster weight (WOW), extracted oyster fresh meat weight (FMW), anterior–posterior length (APL), dorsal–ventral height (DVH), and shell depth (ShD), were manually recorded from each oyster in the field using ruler (mm) or a digital balance (grams).

RGB images were captured under standardised lighting using an XCD LED Portable Photo Studio and a Sony Alpha A6400 mirrorless camera. Each oyster was photographed from the upper-shell, lower-shell and half-shell views. A ruler and an X-Rite ColorChecker Passport were included in each image to support size calibration and colour standardisation. After quality control, 697 unique oysters with complete image and manually-measured phenotype records were retained for feature extraction and machine learning AI analyses. Whole oyster weight (TOW) and fresh meat weight (FMW) were used as the target variables for machine learning model development, whereas image-derived anterior–posterior length (APL) and dorsal–ventral height (DVH) were independently validated against manually measured image-based reference values to assess the accuracy of the automated shell morphometric measurements (**Figure 1A, B**).

All image processing and machine learning algorithms were implemented in Python 3.12.3 (Van Rossum & Drake, 1995). Oyster segmentation was performed using Ultralytics YOLOv8 (v8.3.28) with PyTorch 2.5.1 (CUDA 12.1) (Paszke et al., 2019) for model training and inference. Image processing and feature extraction were conducted using OpenCV 4.10.0 (Team, 2025). Image annotation and dataset management were performed using the Roboflow web platform. Machine learning models were developed using Scikit-learn 1.6.1, while XGBoost 3.2.0 (Chen & Guestrin, 2016) and CatBoost 1.2.10 (Prokhorenkova et al., 2018) were used for gradient boosting-based regression models.

To enable automated AI image analysis, oyster images were first manually annotated by a human operator in Roboflow. The human annotator produced a segmentation mask around the target areas. These annotations were then used to train AI models (**Figure 1C, D**). Two YOLOv8 instance segmentation models were trained: one to segment upper- and lower-shell images and another to segment meat and gonad regions from half-shell images. The resulting trained models were then used to generate segmentation masks on all oysters to automatically extract morphological features from shell and tissue images.

To validate the automated morphometric measurements, a custom software application was developed to manually measure APL and DVH from upper-shell images using the same anatomical definitions as conventional manual measurements (**Figure 1E, F**). Because shell orientation varied during image acquisition, an automated hinge alignment algorithm was developed to standardise shell orientation before feature extraction (**Figure 1G, H**). Following alignment, AI-derived APL and DVH were automatically extracted from the segmented shell masks using the same anatomical definitions as the manual image-based measurements.

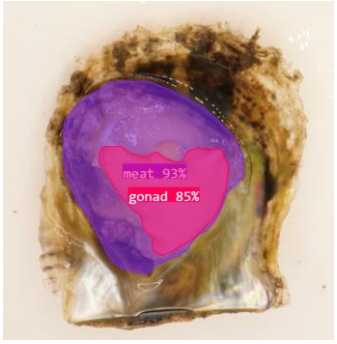
Morphological features extracted from the segmentation masks included shell area, shell perimeter, meat area, gonad area, total tissue area, AI-derived APL, and AI-derived DVH. Pixel-based measurements were converted to physical units using image calibration factors and combined into a feature matrix for subsequent machine learning analyses.

(A)

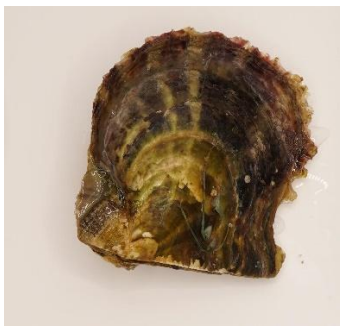
(B)

(C)

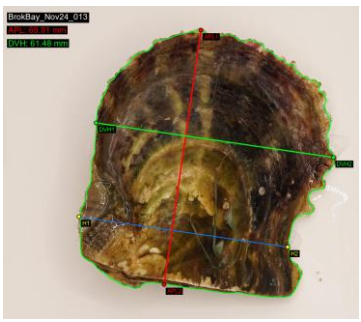
(D)



(E)



(F)



(G)

(H)

Figure 1 Image annotation, morphometric measurement, and hinge alignment workflow for AI-based computer-vision image analysis. **(A)** Original upper-shell image. (**B)** Standardised upper-shell image after automatic hinge alignment. **(C)** Original half-shell image. **(D)** Roboflow-annotated half-shell image showing the segmented meat and gonad regions. **(E)** Original upper-shell image used for manual image-based measurement of anterior–posterior length (APL) and dorso–ventral height (DVH). **(F)** Manual identification of the hinge line (blue), with APL (red) and DVH (green) measured using the same anatomical definitions as the manual shell measurements. **(G)** Original upper-shell image prior to automated hinge alignment. **(H)** Standardised upper-shell image after automatic hinge alignment.

Objective 2: Develop near-field infrared (NIR) prediction models to estimate oyster meat quality traits, including glycogen and lipid content

NIR data acquisition

Data training to evaluate NIR as a technology to estimate glycogen and lipid content of oyster muscle was based on oysters from two sampling batches: June 2025 (n = 150) and December 2025 (n = 414). Here, newly harvested and fresh oysters were scanned on-site at Broken Bay Pearl Farm using a MicroNIR™ 1700 EC spectrometer (VIAVI Solutions Inc., USA) after cleaning, shucking, and thoroughly removing excess surface moisture to improve spectral quality (**Figure 2**). In the December 2025 oyster batch sampled NIR spectra were collected from the gonad, adductor muscle, mantle, and remaining soft tissues of each oyster, while in June 2025 oysters, only the adductor muscle was scanned due to gonads being poorly developed outside the spawning season.

Initial analyses of the June 2025 data showed that the high moisture content of fresh oyster tissue reduced spectral quality and limited prediction accuracy. Consequently, the NIR sampling protocol was modified for the December 2025 batch to include freeze-drying the oyster tissue before NIR analysis to minimise moisture interference.

For the freeze-dried protocol, whole oyster tissues were processed and frozen on the day of harvest and transported to JCU. Samples were stored frozen until lyophilisation for 96 h using a Christ Alpha 2-4 freeze dryer (Martin Christ). The dried tissue was then ground into a fine powder using a ø 25 mm tissue pestle attached to a power drill and scanned in triplicate using the same NIR instrument (**Figure 2**). Following spectral acquisition, samples were retained for biochemical analyses of glycogen and lipid content, which were used as reference values for calibration and validation of the NIR prediction models.

Figure 2 Overview of NIR spectral acquisition and sample preparation methods used in this study. (Top left) VIAVI MicroNIR 1700 EC portable near-infrared spectrometer used for all spectral acquisitions. (Bottom left) Shucked Akoya oyster showing the tissue regions scanned: adductor muscle (red), gonad and digestive gland (yellow), and mantle (green). (Centre) Solarbio Glycogen Assay Kit (Solarbio Life Sciences, Beijing, China) used for colourimetric glycogen quantification via the anthrone method to generate reference values for model calibration. (Top right) 25 mm tissue homogeniser pestle fitted to a power drill, used for mechanical grinding of freeze-dried oyster tissue prior to spectral acquisition. (Bottom right) NIR spectral scan of powdered freeze-dried Akoya oyster tissue acquired using the MicroNIR 1700 EC.

Biochemical analysis of glycogen and lipid content

Total Moisture

Whole-oyster moisture content was determined by comparing fresh and freeze-dried weights following lyophilisation.

$$Moisture\ (\%) = \frac{\text{Fresh weight} - \text{Dry weight}}{\text{Fresh weight}} \times 100$$

Crude lipid concentration

Crude lipid was quantified from freeze-dried oyster tissue using a modified solvent extraction method based on the protocols of Folch et al. (1957) and Bligh and Dyer (1959). Briefly, 100 mg of freeze-dried tissue was extracted with dichloromethane:methanol (2:1),

followed by phase separation, solvent evaporation, and gravimetric determination of lipid content. Lipid concentration was calculated as:

$$Lipid\ concentration\ (mg\ g^{-1}\ DMW) = \frac{lipid\ mass\ (g)}{sample\ mass\ (g)} \times 1000$$

Where DMW denotes dry meat weight. Method repeatability was assessed using 10 samples analysed in triplicate, resulting in an average relative error of approximately 8%.

Glycogen concentration

Glycogen concentration was quantified using a commercial glycogen assay kit (Solarbio, BC0345, China), following the manufacturer's colourimetric protocol with minor modifications (**Figure 2**). Adductor muscle was analysed for the June 2025 samples, whereas whole-oyster tissue was analysed for the December 2025 samples. Briefly, 10 mg of homogenised tissue was digested in potassium hydroxide, reacted with the anthrone reagent, and glycogen concentration was determined colourimetrically by measuring absorbance at 620 nm using a Spectrostar Nano spectrophotometer (BMG Labtech). Glycogen concentration was calculated as:

$$Glycogen\ concentration\ (mg\ g^{-1}\ DMW) = \frac{4.50 \times (OD_{standard}/OD_{sample})}{sample\ mass\ (g)}$$

Spectral signal preprocessing and chemometric modelling

Prior to multivariate analysis, spectral outliers were identified and removed using a principal component analysis (PCA)–Mahalanobis distance approach (MD > 3). The remaining samples were randomly divided into calibration (75%) and validation (25%) datasets.

For the June 2025 samples, multiple spectral preprocessing methods and machine learning algorithms were evaluated to optimise prediction accuracy. Scatter correction methods trialled included the wavelet prism extension algorithm (Chen et al., 2004), multiplicative scatter correction (MSC), standard normal variate (SNV), and minimum redundancy-maximum relevance (mRMR) wavelength selection (Ma et al., 2024). Based on these results,

a streamlined preprocessing pipeline was developed and applied to the December 2025 samples to improve computational efficiency (**Figure 3**).

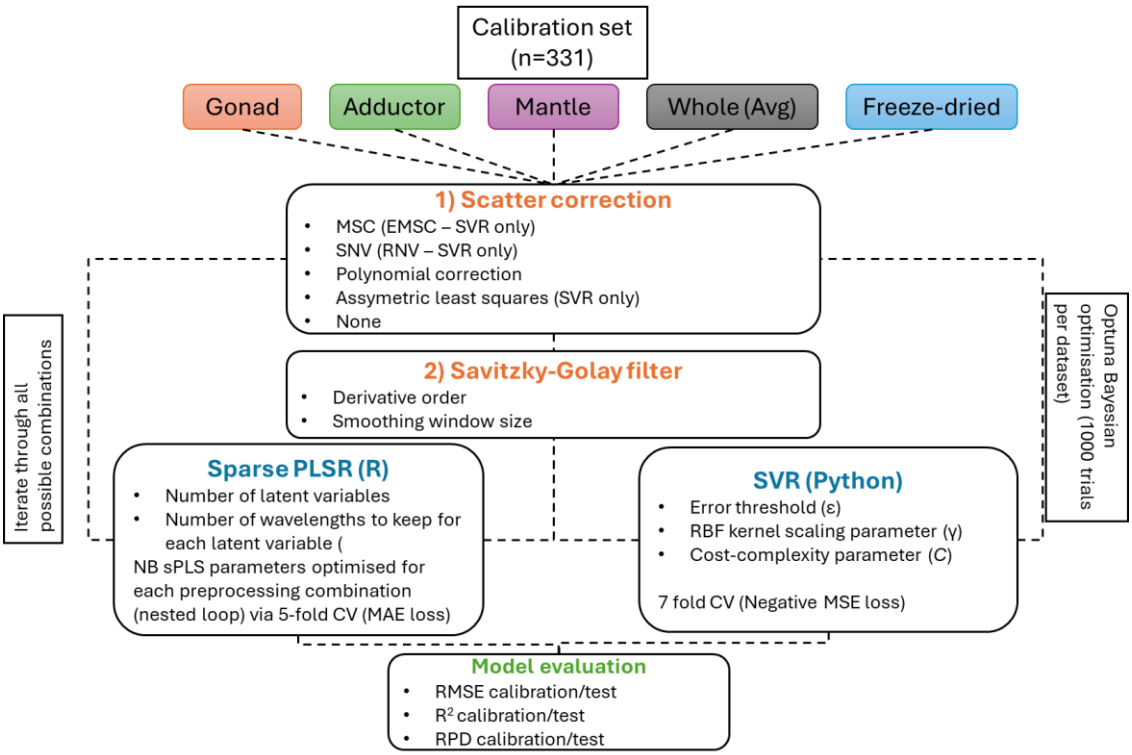


Figure 3 The streamlined pipeline developed to navigate this combinatorial search space efficiently.

Three predictive models were evaluated: sparse partial least-squares (sPLS), support vector regression (SVR), and a hybrid convolutional neural network–long short-term memory (CNN–LSTM) model. The first three are established chemometric approaches, while CNN-LSTM represents an emerging deep learning method increasingly applied to spectral modelling. All analyses were performed using Python v3.11.

Objective 3 Develop and evaluate a custom target SNP panel to support genetic improvement in the Australian Akoya oyster

A total of 234 Australian Akoya oysters from New South Wales (NSW) and Western Australia (WA) were used for custom SNP panel development. Panel performance was validated using 190 samples representing 178 unique oysters from the two populations. Genomic DNA (gDNA) was obtained from non-invasive swab samples or foot muscle tissue collected as part of the current and previous JCU projects. DNA was extracted using either modified CTAB or

automated Chemagic™ extraction protocols. Extracted gDNA were submitted to the Australian Genome Research Facility (AGRF) for library preparation and sequencing.

SNP discovery and custom panel development

Double-digest restriction-site associated DNA sequencing (ddRAD-seq) SNP libraries were generated for 214 Australian Akoya oysters, while an additional 20 individuals were subjected to low-pass whole-genome sequencing (WGS) (**Figure 4**). Sequence reads from both ddRAD and WGS individuals were aligned to the Akoya oyster (*Pinctada fucata martensii*) reference genome (ASM3311930v1; GenBank: GCA_033119305.1; (Du et al., 2017; Zheng, 2023)), and SNPs identified using using a custom Nextflow workflow, Marine Omics Variant Pipeline (MOVP; https://github.com/marine-omics/movp/blob/main/README.md). After quality filtering, 341,239, 4,017, and 2,191 SNPs were retained from the WGS, ddRAD-NSW, and ddRAD-WA datasets, respectively, and were retained for subsequent downstream SNP selection for panel development (**Figure 4**).

Candidate SNPs for the genotyping panel were prioritised to support key breeding applications, including parentage assignment, population genetic analyses, and genomic selection. SNPs for the panel were chosen to be evenly distributed across the genome to maximise genome coverage and minimise gaps between SNP markers. Additional 200 SNPs putatively associated with growth and nacre colour were included on the panel based on published genome-wide association studies (Wang et al., 2022; Zheng et al., 2023). This statistical process produced 11,999 high-quality candidate SNPs for custom panel design, of which 9,448 were successfully incorporated into the final Allegro SNP panel, providing genome-wide coverage across all 14 chromosomes of Akoya oyster.

Performance validation of custom target SNP panel

The performance of the custom target SNP panel was evaluated at the sample and SNP levels using 190 samples, comprising 178 unique Akoya oysters and 12 technical duplicate samples included to assess genotyping reproducibility. Oyster sample-level performance was assessed using sequencing read depth, SNP call rate, and proportion of missing SNP genotypes, whereas SNP-level performance was evaluated based on individual SNPs call rate, missing genotype rate, and the number of polymorphic SNPs.

Accuracy was estimated by comparing panel genotypes with WGS data from 19 individuals. Repeatability was assessed using six oysters that were genotyped twice within the same sequencing batch, while reproducibility was evaluated using six individuals genotyped independently across two sequencing batches. Concordance of SNP calls were calculated at both the sample and SNP levels as the proportion of matching genotype calls between duplicate oysters genotyped.

Further details of SNP discovery, panel development, validation procedures, and quality control are provided in Supplementary Technical Report S1.

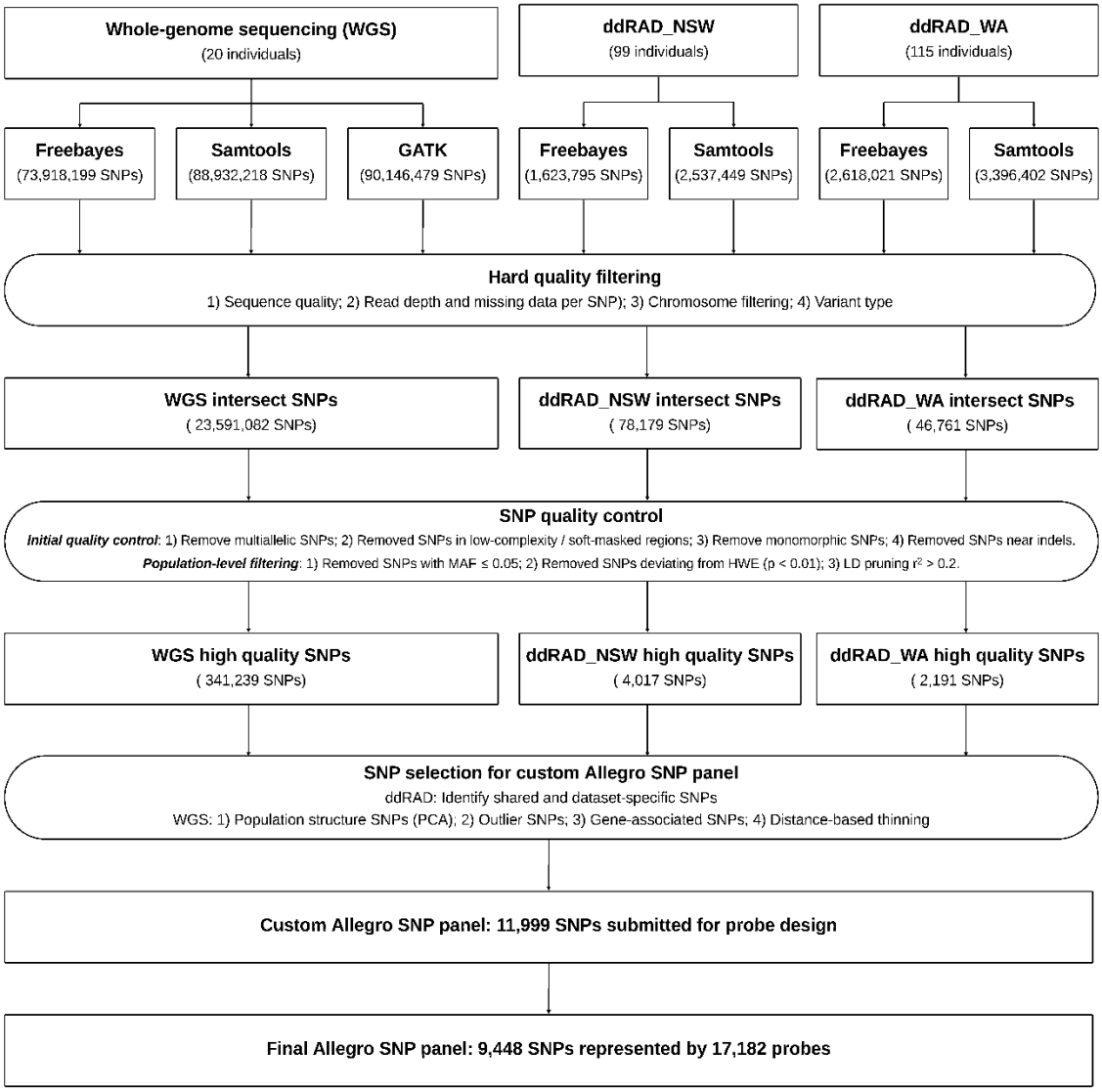


Figure 4 Workflow outlining the steps used for SNP calling, filtering, and selection of SNPs for the custom target SNP panel.

Objective 4: Optimize genomic prediction models for predicting breeding values of growth and meat quality traits

Sample collection, DNA extraction, and SNP genotyping

A total of 803 Akoya oysters from Broken Bay Pearl Farm were sampled across three collection periods: November 2024 (n = 152), June 2025 (n = 140), and December 2025 (n = 511). According to farm records, these oysters originated from three mass-spawning events conducted in October 2021, March 2022, and November 2022. As pedigree records were unavailable and oysters from different spawning events had been mixed during routine farm

operations, the population was treated as a mixed breeding population for genomic analyses. Sampling batch was included in downstream analyses where appropriate.

Genomic DNA (gDNA) was extracted from oyster foot muscle tissue using the Chemagic™ Tissue 10 mg DNA Extraction Kit (Revvity) on a Chemagic™ 360 automated workstation (Revvity), following the manufacturer's instructions and then submitted to the Australian Genomic Research Facility (AGRF) for library preparation and genotyping using the custom target Allegro SNP panel developed in objective 3.

803 Akoya oysters were genotyped using the custom target SNP panel, generating genotype data from each oyster for 9,143 SNP loci. Standard quality control filtering was applied using BCFtools v1.20 (Li, 2011) and VCFtools v0.1.16 (Danecek et al., 2011), including filters for genotype quality, SNP call rate ≥ 90%, MAF ≥ 0.04. Following quality control, 4,970 high-quality SNPs with a sample call rate exceeding 96% were retained for parentage assignment, heritability estimation, and genomic prediction analyses.

Determining phenotypic variation and correlation

Phenotypic data were collected from 803 Australian Akoya oysters for five growth traits: whole oyster weight (WOW), fresh meat weight (FMW), anterior–posterior length (APL), dorsal–ventral height (DVH), and shell depth (ShD). All oysters were cleaned prior to measurement. Shell dimensions were measured using standard morphometric methods, and oyster and meat weights were recorded using a digital balance. Trait distributions were assessed prior to genomic analyses.

Meat quality phenotypes were measured from 414 Australian Akoya oysters collected during the December 2025 sampling. Whole-oyster tissue was freeze-dried, homogenised, and used for biochemical analyses. The phenotypes included dry meat weight (DMW), glycogen concentration (GC; mg g^{-1} DMW), lipid concentration (LC; mg g^{-1} DMW), total glycogen content (TGC; mg per oyster), and total lipid content (TLC; mg per oyster). Total glycogen and lipid contents were calculated by multiplying their respective concentrations by whole-oyster dry meat weight.

Parentage assignment and pedigree reconstruction

Parentage assignment and pedigree reconstruction were conducted using 949 SNPs to characterise the family structure of 803 farm oysters as putative offspring with all available candidate broodstock genotypes. Parent–progeny relationships were inferred using CERVUS v3.0.7 (Kalinowski et al., 2007) and COLONY v2.0.7.0 (Jones & Wang, 2010), following the approach described by Vu et al. (2026). Parentage assignments supported by both programs at the 95% confidence level were accepted. Assignments identified by only one program (achieved 95% confidence) were further evaluated using genomic relationships estimated from a genomic relationship matrix (GRM) calculated with the snpgdsGRM function in the R package SNPRelate v1.42.0 (Zheng et al., 2012).

Heritability and genetic correlation estimation

Heritability and genetic correlations for growth and meat quality traits were estimated to evaluate the genetic architecture of economically important traits and identify suitable targets for selective breeding. Genomic best linear unbiased prediction (GBLUP) and pedigree-based best linear unbiased prediction (PBLUP) models were used to estimate trait heritability, while genetic and phenotypic correlations among traits were estimated using bivariate GBLUP models. These analyses were implemented in ASREML v4.2 (Butler et al., 2017). Growth traits were analysed with sampling batch fitted as a fixed effect, whereas meat quality traits were analysed without a batch effect. Genomic analyses were based on a genomic relationship matrix constructed from 4,970 SNPs, while pedigree analyses used the reconstructed pedigree generated in this study.

Genomic prediction analyses

Genomic prediction analyses were conducted to estimate genomic estimated breeding values (GEBVs) for five growth traits and five meat quality traits using 4,970 high-quality SNPs. Growth trait analyses were performed on 803 Australian Akoya oysters collected across three sampling batches, while meat quality analyses were conducted on a subset of 414 oysters collected during the December 2025 sampling.

Six genomic prediction methods representing linear mixed models, Bayesian, kernel-based, and machine-learning approaches were evaluated, including genomic BLUP (GBLUP), pedigree-based BLUP (PBLUP), BayesC, multiple-kernel reproducing kernel Hilbert spaces regression (mRKHS), Random Forest (RF), and Extreme Gradient Boosting (XGBoost). GBLUP and PBLUP were implemented in ASReml v4.2 (Butler et al., 2017), BayesC and mRKHS in the R package BGLR v1.1.2 (Pérez & de Los Campos, 2014), and RF and XGBoost (Chen & Guestrin, 2016)using the R package randomForest 4.7-1.1 (Liaw & Wiener, 2002) and xgboost v3.3.1.1 (Chen et al., 2026), respectively. Model performance was evaluated using repeated five-fold cross-validation. Prediction ability and prediction accuracy were compared among models to identify the most suitable approach for genomic selection in Akoya oysters.

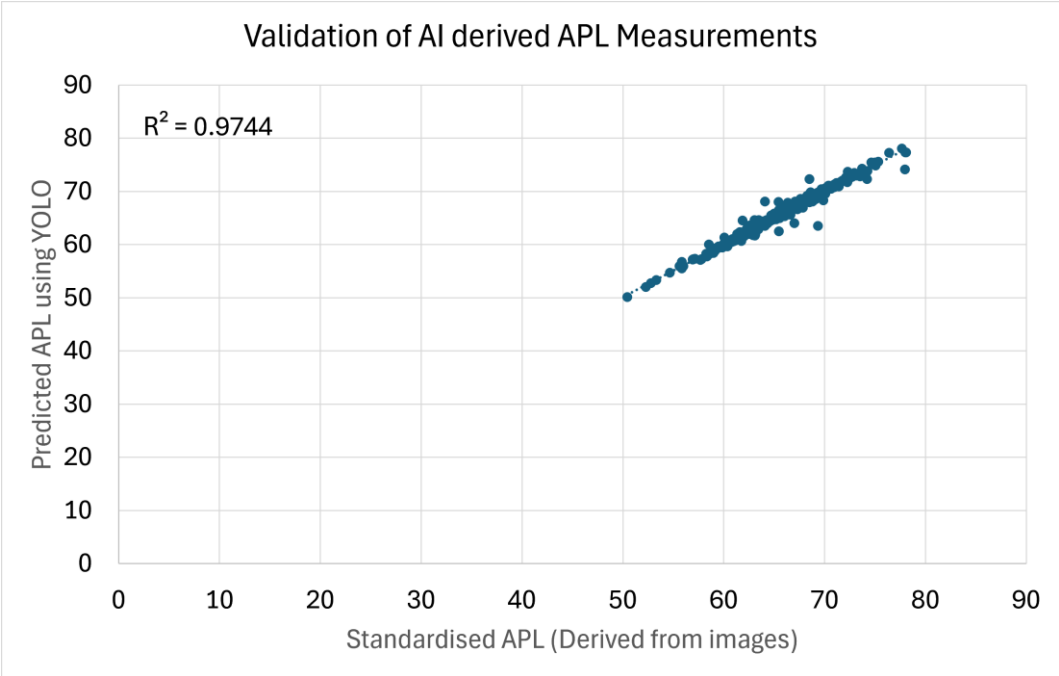
RESULTS

Project objective 1:

Validation of AI-derived shell morphometric measurements

To evaluate the accuracy of the proposed YOLO-based automated morphometric extraction framework, AI-derived anterior–posterior length (APL) and dorsal–ventral height (DVH) measurements were compared with manually measured reference values obtained from standardised shell images, instead of human-measured APL and DVH on the actual oyster shell. Manual measurement of APL and DVH requires both dimensions to be defined relative to the oyster hinge, which can introduce inconsistencies when shell orientation varies between images. To standardise this process, all images from the November 2024 and June 2025 datasets were first aligned with the oyster hinge positioned at the bottom of the image. APL and DVH were then manually measured using custom measurement software, ensuring that the dimensions were determined accurately, consistently, and in accordance with the intended manual human measurement procedure. These measurements were used as the reference standard for validation. The AI-derived measurements showed strong agreement with the reference measurements (APL: R^2 = 0.97, **Figure 5A**; DVH: R^2 = 0.95, **Figure 5B**), demonstrating the accuracy and reliability of the proposed framework.

(A)



(B)

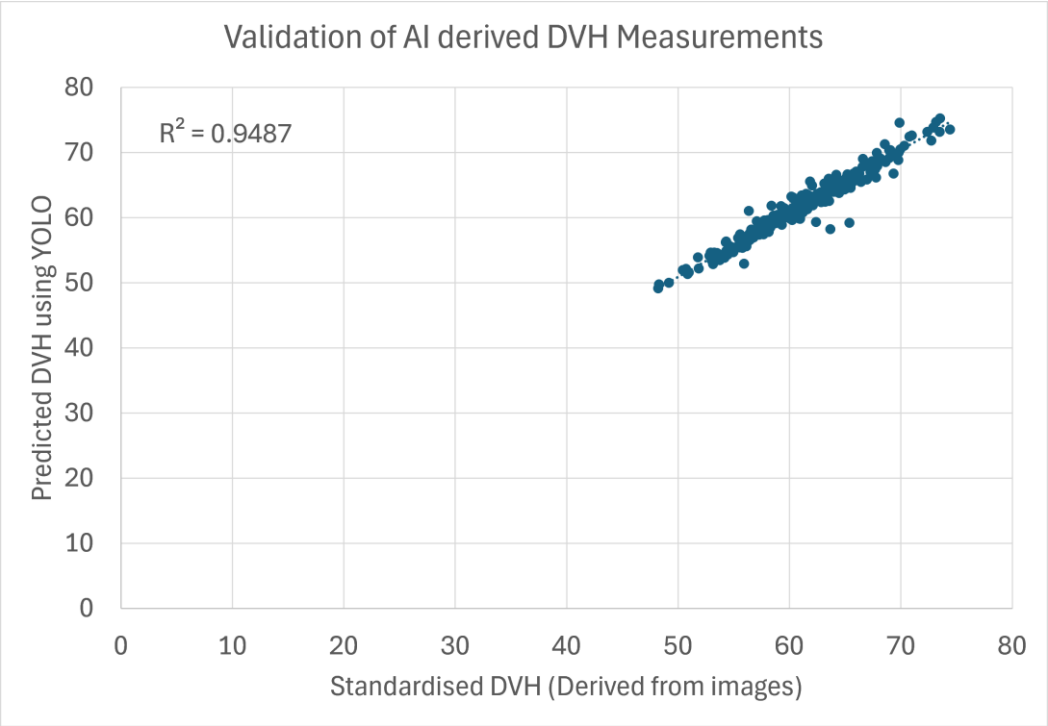


Figure 5 Validation of AI-derived shell morphometric measurements. **(A)** Comparison between manually measured image-based anterior–posterior length (APL) and AI-derived APL obtained using the automated hinge alignment and morphometric extraction framework (R^2 = 0.9744). **(B)** Comparison between manually measured image-based dorso–ventral height (DVH) and AI-derived DVH obtained using the same framework (R^2 = 0.948).

Whole Oyster Weight Prediction

For whole-oyster weight prediction, Ridge Regression was the best-performing model, achieving an R^2 of 0.83 for the overall dataset, a mean cross-validation R^2 of 0.81 ± 0.03, and a test-set R^2 of 0.79. On the test set, the model achieved a Mean Absolute Error (MAE) of 2.87 g and a Root Mean Square Error (RMSE) of 3.87 g, indicating that its predictions typically differed from the measured oyster weights by approximately 3 g. MAE gives the average absolute prediction error and is easy to interpret: predictions were off by 2.87 g on average. RMSE penalises large errors more strongly because errors are squared before averaging. It therefore indicates whether occasional predictions were substantially less accurate.

Figure 6 shows the relationship between the measured and predicted whole oyster weights obtained using Ridge Regression. The high linear correlation indicates that shell morphological features extracted from the standardised images provide reliable predictors of whole oyster weight while maintaining good generalisation performance.

Although ensemble algorithms achieved substantially higher training accuracies (>0.96), they exhibited larger reductions in testing performance, suggesting overfitting. The success of Ridge Regression indicates that whole oyster weight is predominantly governed by linear relationships between shell morphology and overall body size, making regularised linear regression more appropriate than highly flexible nonlinear ensemble models.

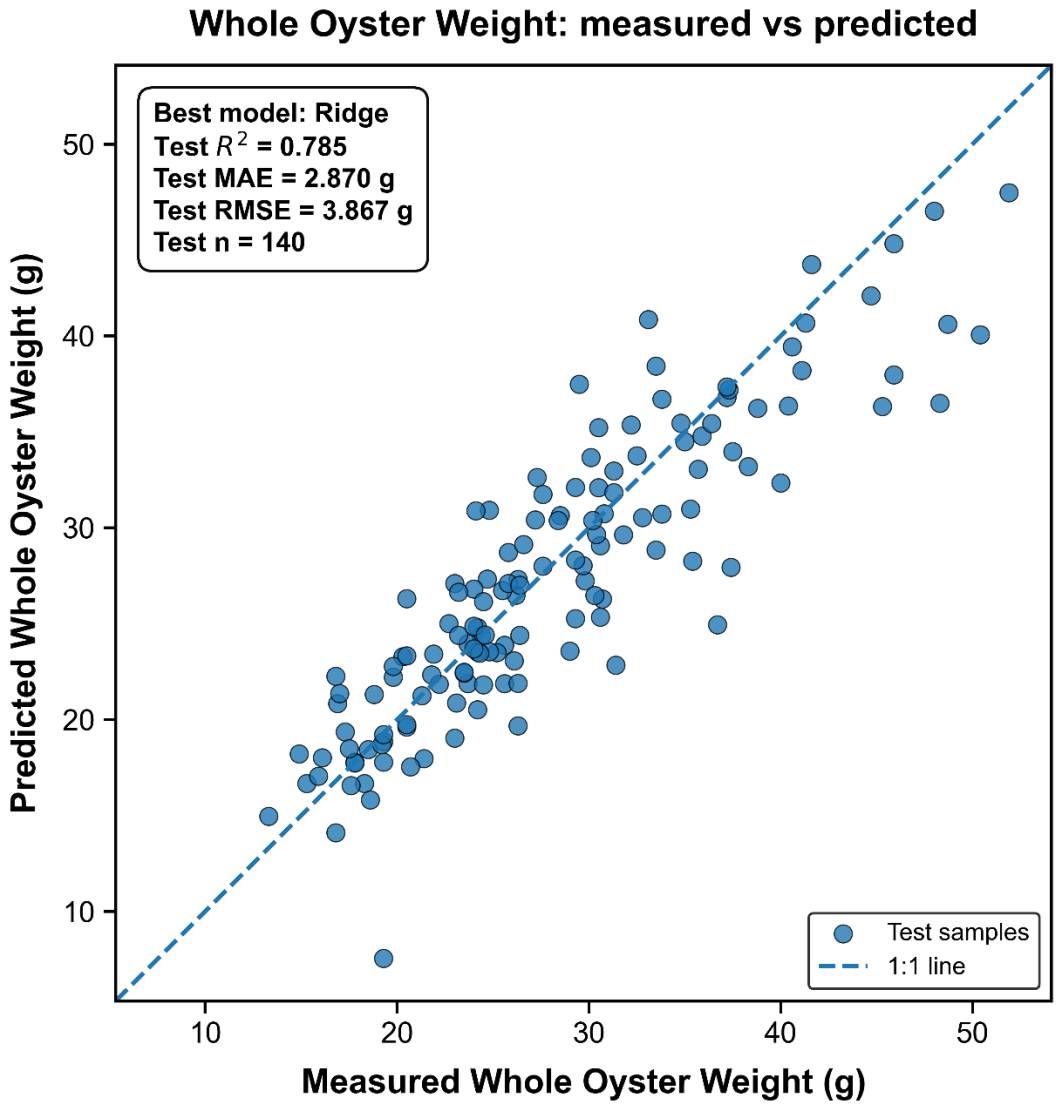


Figure 6 Comparison between measured and predicted whole oyster weight obtained using the Ridge Regression model.

Oyster Meat Weight Prediction

Out of several machine learning models developed, Support Vector Regression (SVR) showed the best overall performance for meat weight estimation, with a Training R^2 = 0.90, Cross-

Validation R^2 = 0.86 ± 0.0212, and Test R^2 = 0.84 **(Figure 7)**. Prediction errors were low, with an MAE of 1.06 g and an RMSE of 1.40 g, indicating high generalisation to previously unseen samples.

The relationship between the measured and predicted meat weights is illustrated in **Figure 7**. The close agreement between the observed data and the fitted regression line demonstrates the ability of the proposed SVR model to accurately estimate oyster meat weight from the extracted image-derived morphological features.

Tree-based ensemble models, including Random Forest, Extra Trees, Gradient Boosting, XGBoost, and CatBoost, achieved very high training accuracies, with some exceeding Training R^2 = 0.99. However, their lower testing performance indicated varying degrees of overfitting. The relatively small difference between the training and testing performance of SVR demonstrates its superior ability to model the nonlinear relationship between tissue morphology and meat biomass while maintaining good generalisation.

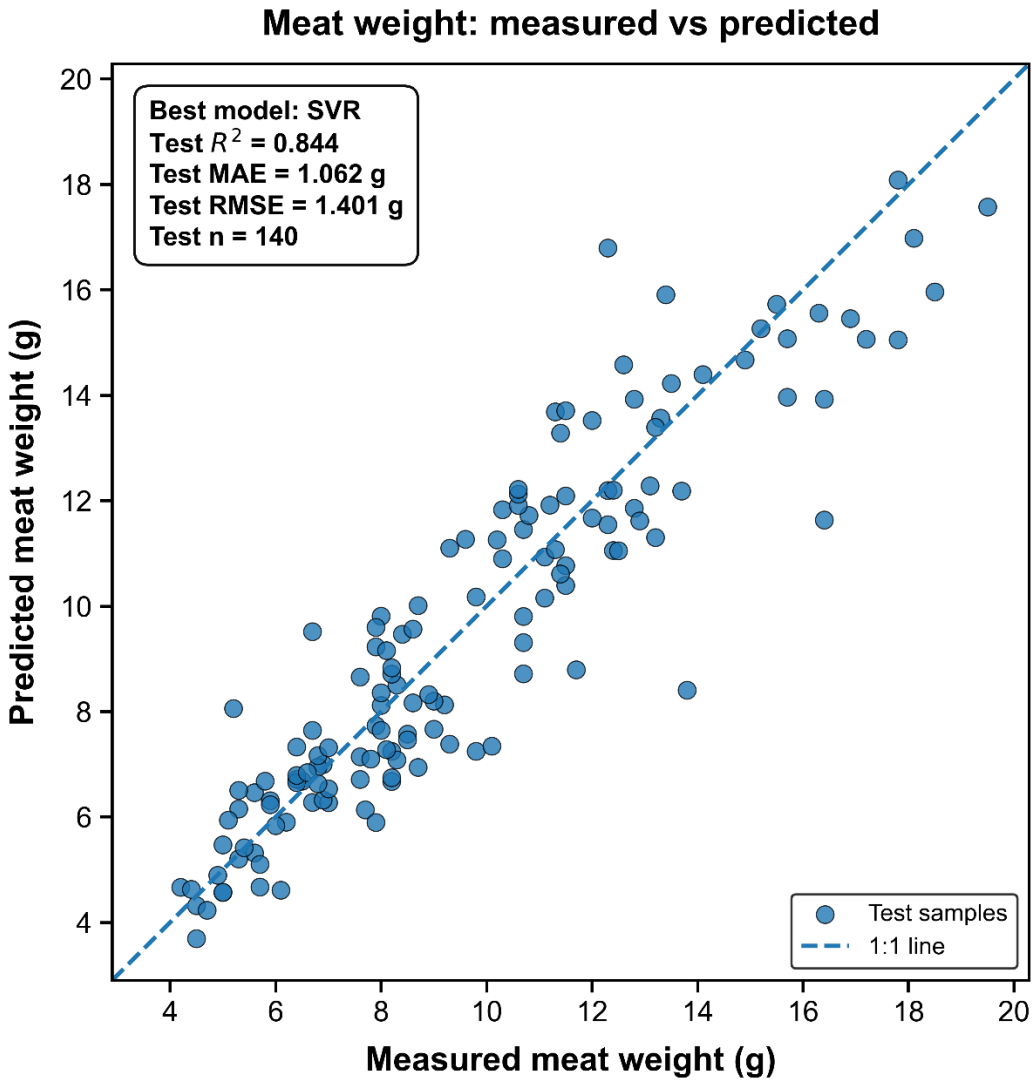


Figure 7 Comparison between measured and predicted meat weight obtained using the Support Vector Regression (SVR) model.

Project objective 2:

The heterogeneous surface morphology of fresh Akoya oyster tissue, combined with its inherently high moisture content (mean 84% wet weight, range 61.2–94.2%), presented substantial challenges for NIR spectral acquisition. Surface irregularities increase light scattering and reduce the intensity of diffuse-reflected light, while the strong and broad O-H absorption bands associated with free water attenuate and mask the weaker absorbance signatures of target analytes such as glycogen and lipid. Together, these effects introduce high-frequency, sample-specific spectral distortions that are difficult to correct through preprocessing alone. The consequence was low and highly variable signal intensity across the spectral range, which translated to poor predictive performance across all chemometric models evaluated. For fresh samples, R^2 of the test set ranged from 0.29 to 0.58 for glycogen and 0.004 to 0.60 for lipid (**Table 1**), indicating that models built on fresh tissue spectra lacked sufficient discriminatory power for reliable quantitative prediction.

Table 1 R^2 values of the test set, derived from regression of predicted against observed values for glycogen and lipid across all chemometric models. For glycogen, the highest R^2 achieved from NIR scans of fresh tissue was 0.58 using sparse partial least-squares regression (sPLS), which improved to 0.74 when spectra were acquired from freeze-dried samples. For lipid, model performance remained poor across all sample preparation methods and model types.

Analyte	Tissue	sPLS	SVR	CNN-LSTM
GLYCOGEN	GONAD	0.337	0.338	0.287
	ADDUCTOR	0.500	0.455	0.461
	MANTLE	0.345	0.464	0.357
	WHOLE BODY	0.584	0.576	0.507
	FREEZE DRIED	0.709	0.736	0.649
LIPID	GONAD	0.527	0.580	0.603
	ADDUCTOR	0.004	0.046	0.027
	MANTLE	0.110	0.162	0.205
	WHOLE BODY	0.509	0.583	0.501
	FREEZE DRIED	0.551	0.566	0.498

Freeze-drying the December 2025 samples prior to spectral acquisition substantially improved model performance. Following moisture removal, R^2 for glycogen increased to 0.74 with an RPD of 1.948, indicating acceptable predictive ability and a meaningful improvement in signal quality. This finding is consistent with the broader NIR literature on bivalve species, where freeze-drying prior to scanning is consistently reported as a prerequisite for reliable chemometric modelling. Brown et al. (2012), Wang et al. (2015), and Fu et al. (2024), all working with the Pacific oyster *Crassostrea gigas*, achieved strong predictive models

exclusively using freeze-dried, homogenised tissue, with R^2 values exceeding 0.90 for glycogen.

The comparatively modest R^2 values obtained in the present study likely reflect a combination of factors, including tissue heterogeneity inherent to Akoya oyster, instrument characteristics of the portable MicroNIR 1700 EC, and the reference chemical analysis used to derive glycogen and lipid values.

One probable contributing factor is the morphological distinctiveness of Akoya oyster relative to better-studied species such as *Crassostrea gigas* and *Saccostrea glomerata*, for which most existing NIR calibration models have been developed. In Pacific oysters, the digestive gland and gonad-surrounding mantle area together account for over 90% of total tissue mass, with approximately 47% and 44% respectively, with the adductor muscle comprising only ~9% (Berthelin et al., 2000). This relatively consistent tissue architecture means that a surface scan captures a broadly representative and reproducible biochemical profile across individuals. In contrast, the body surface composition of Akoya oyster varies considerably among individuals and across seasons (**Figure 8**). During the spawning season, the visceral mass is consist of gonad and digestive gland, which dominates the exposed tissue surface following shucking, while the adductor muscle occupies a comparatively smaller area. Outside the spawning season, however, there is substantial inter-individual variation: some individuals present a larger gonad and digestive gland surface area, while others are dominated by mantle tissue or adductor muscle. This morphological variability means that a fixed-position surface scan captures a biochemically inconsistent snapshot across individuals, introducing systematic noise into spectral readings that cannot be attributed to analyte concentration alone.

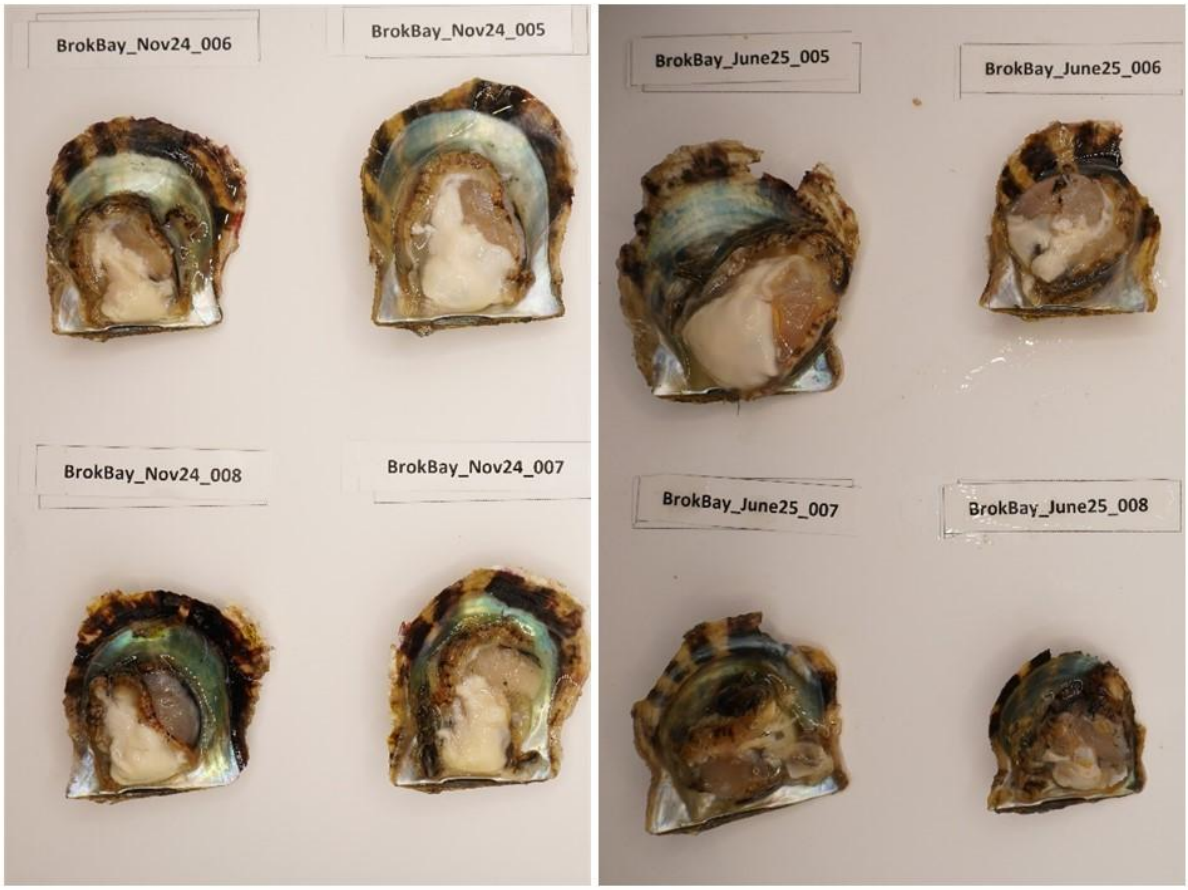


Figure 8 Shucked Akoya oyster (*Pinctada fucata*) samples collected during the spawning season (left) and outside the spawning season (right). During the spawning season, the gonad is comparatively enlarged and visually prominent, dominating the exposed tissue surface. Outside the spawning season, considerable morphological variability is evident among individuals in the relative surface area occupied by the gonad, mantle, and adductor muscle.

This heterogeneity is compounded by the uneven spatial distribution of glycogen and lipid across tissue types. Glycogen and lipid are concentrated primarily in the digestive gland and gonad, while the adductor muscle contains markedly lower concentrations of both analytes, with approximately 0.59% glycogen and 1.17% lipid on a wet weight basis (Berthelin et al., 2000; Min et al., 2020). Consequently, a scan acquired from any single tissue region in isolation is a poor predictor of whole-body glycogen or lipid content, as it reflects only the local biochemical composition of that region rather than the integrated nutritional status of the animal. This is supported by the present results, which demonstrate that whole-body scans (averaging spectral readings across the three tissue positions) consistently outperformed single-tissue scans in predictive power for both glycogen and lipid (**Table 1**).

Instrumental differences between studies represent a further source of variation in model performance. In the present study, a portable VIAVI MicroNIR 1700 EC spectrometer was selected to facilitate integration of NIR readings into routine aquaculture farm operations, offering a practical and cost-effective alternative to laboratory-grade benchtop instruments. However, this portability comes at the cost of spectral coverage, the MicroNIR 1700 EC

operates over a wavelength range of approximately 908 - 1,676 nm (11,012 - 5,966 cm^{-1}), which is considerably narrower than the ranges employed in comparable studies **Figure 9**.

Li et al. (2023), working with Crassostrea gigas, used a Fourier Transform NIR spectrometer (Thermo Fisher Scientific, USA) scanning 32 times across a wavelength range of 833-2,630 nm. This broader range encompasses several analytically important spectral regions that fall outside the MicroNIR 1700 EC's coverage, including the glycogen combination bands at 1,538-1,887 nm, fatty acid first overtone regions at 1,422-1,791 nm (7,033-5,583 cm^{-1}) and 1,687-1,795 nm (5,926-5,571 cm^{-1}), and the C-H combination bands associated with lipid at 2,301-2,343 nm (4,346-4,269 cm^{-1}). The last of these regions — around 2,300 nm — falls entirely outside the MicroNIR 1700 EC's detectable range, representing a meaningful loss of chemometric information for lipid prediction in particular.

Similarly, Brown et al. (2012) used an Analytical Spectral Devices LabSpec 5000 spectrometer covering the full 350–2,500 nm (28,571–4,000 cm^{-1}) range in their study of Sydney rock oyster (*Saccostrea glomerata*) and Pacific oyster (*Crassostrea gigas*). This instrument captured the key glycogen absorption bands at 1,400–2,300 nm and fat-related bands at 1,100–1,700 nm in their entirety. The substantially wider spectral window available to these instruments likely contributed to their superior model performance relative to the present study, as they retained full access to all major NIR absorption features associated with the target analytes.

Fu et al. (2024) employed the MicroNIR 1700 (908–1,676 nm) to estimate glycogen in *C. gigas*. They reported R^2 values exceeding 0.90 for glycogen prediction. However, this study is the absence of any wavelength-level interpretability analysis, thus they did not identify or discuss which specific spectral regions drove glycogen prediction in their models, instead citing prior literature for band assignments rather than conducting variable importance analysis on their own data. Without knowing which wavelength regions were informative for glycogen concentration in *C.gigas*, it is difficult to determine whether the MicroNIR 1700 EC used in the present study captures sufficient spectral information for reliable glycogen prediction, or whether the absence of coverage beyond 1,676 nm represents a critical limitation.

A further source of discrepancy between the present study and comparable NIR literature relates to the reference chemical methods used to derive analyte concentrations for model calibration. Brown et al. (2012), Li et al. (2023), and Fu et al. (2024) all employed relatively sophisticated analytical techniques for glycogen and lipid quantification, which likely contributed to the precision and accuracy of their calibration reference values.

For glycogen quantification, the present study employed the anthrone colourimetric assay (Carroll et al., 1956), a simple, cost-effective, and widely used method in shellfish and aquaculture research. The assay estimates total carbohydrate content through a colourimetric reaction. However, the anthrone method is non-specific and susceptible to interference from other carbohydrate species present in biological tissue, which can introduce systematic error into reference values. Chromatographic methods such as high-performance liquid chromatography (HPLC) offer considerably greater analytical specificity and accuracy by resolving individual carbohydrate components and minimising matrix interference (Getachew et al., 2019). Since the accuracy of an NIR calibration model is fundamentally bounded by the accuracy of its reference method, any imprecision or bias in the anthrone assay propagates directly into the calibration, inflating residual error and

suppressing model performance. A parallel limitation applies to lipid quantification. The present study used crude lipid extraction, a gravimetric, non-targeted method in which all solvent-extractable compounds are collectively isolated and weighed, encompassing triglycerides, phospholipids, sterols, and waxes without discrimination among lipid classes. This is in contrast to the targeted approach employed in several comparable studies, where gas chromatography (GC) or GC coupled with mass spectrometry (GC-MS) is used to quantify individual fatty acid species, including nutritionally and commercially important polyunsaturated fatty acids such as eicosapentaenoic acid (EPA) and docosahexaenoic acid (DHA).

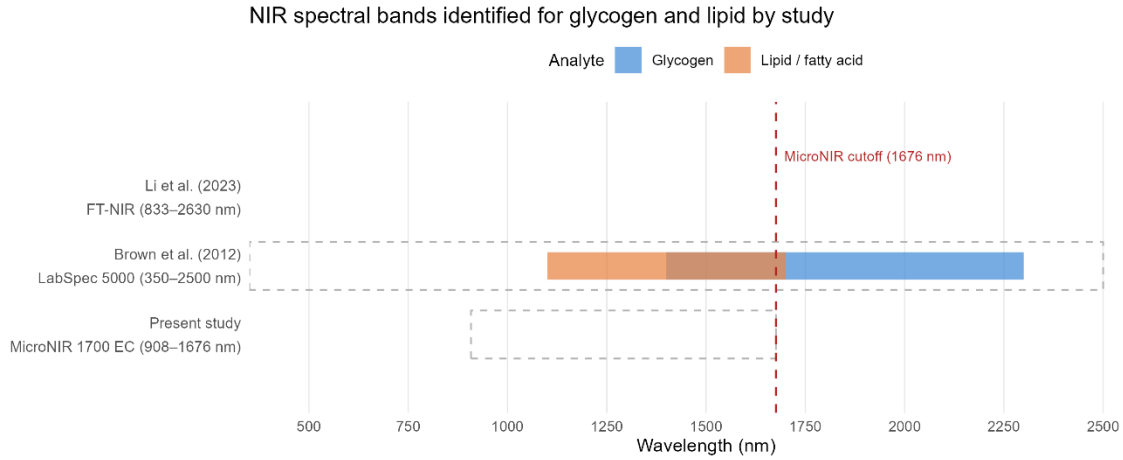


Figure 9 Spectral wavelength ranges of NIR spectrometers used across three studies on Crassostrea gigas, alongside analyte-specific absorption bands identified in each study. Dashed boxes indicate the operational wavelength range of each instrument. Blue bars denote spectral regions associated with glycogen absorption and orange bars denote spectral regions associated with fatty acid absorption, as reported by each respective study. The vertical dashed red line indicates the upper wavelength limit of the MicroNIR 1700 EC (1,676 nm) used in the present study. Note that all band assignments are derived from studies on C. gigas and may not directly translate to Akoya oyster.

Research conclusions and recommendations

This study demonstrates that NIR spectroscopy has the potential to estimate glycogen concentration in Akoya oyster. However, to achieve reasonable predictive performance, spectral acquisition from freeze dried samples is required. This represents a fundamental practical barrier to integration into routine farm operations, where rapid, non-destructive assessment of live or freshly shucked animals is the operational priority. Until a reliable calibration model can be developed for fresh tissue and it may require a benchtop FT-NIR spectrometer with broader spectral coverage, larger and more morphologically diverse training datasets, and more analytically specific reference methods, portable NIR

spectroscopy in its current form cannot be recommended as a standalone condition monitoring tool for akoya pearl oyster aquaculture.

Project objective 3:

This project presented the first bioinformatic pipeline specifically designed for developing and validating a custom SNP genotyping panel for the Australian Akoya oyster using the Allegro targeted sequencing platform, which offers an efficient and cost-effective genotyping strategy. This SNP panel has the potential to accelerate genetic breeding programs aimed at improving economically important traits, support fisheries management, and contribute to the sustainable development of aquaculture.

The custom target Allegro SNP panel was successfully manufactured by Tecan Genomics (Redwood City, CA, USA). Of the 11,999 candidate SNPs submitted for panel design, 9,448 SNPs (78.7%) were incorporated into the final panel (**Figure 10**). Genotype data were successfully generated for 9,041 loci (95.7%) across 190 genotype records representing 178 unique Australian Akoya oysters.

Sample quality was generally high, with an average genotype call rate of 90.4%. Two samples showed poor genotyping performance, likely due to low DNA quantity, and were excluded from subsequent analyses. Following quality control filtering, 5,660 high-quality SNPs from 176 unique individuals were retained for downstream analyses. These SNPs provided genome-wide coverage suitable for parentage assignment, population genetic analyses, and genomic selection.

The custom Allegro SNP panel demonstrated high genotype concordance across all validation analyses (Figure 10). Genotype accuracy, assessed by comparison with whole-genome sequencing (WGS) data from 19 individuals, averaged 91.3%. Repeatability, evaluated using duplicate samples within the same sequencing batch, averaged 97.4%, while reproducibility across two independent sequencing batches averaged 96.6%. These results demonstrate that the custom SNP panel provides accurate, reliable, and robust genotype calls.

Additional quality control metrics and genotype validation results are provided in **Appendices 1**.

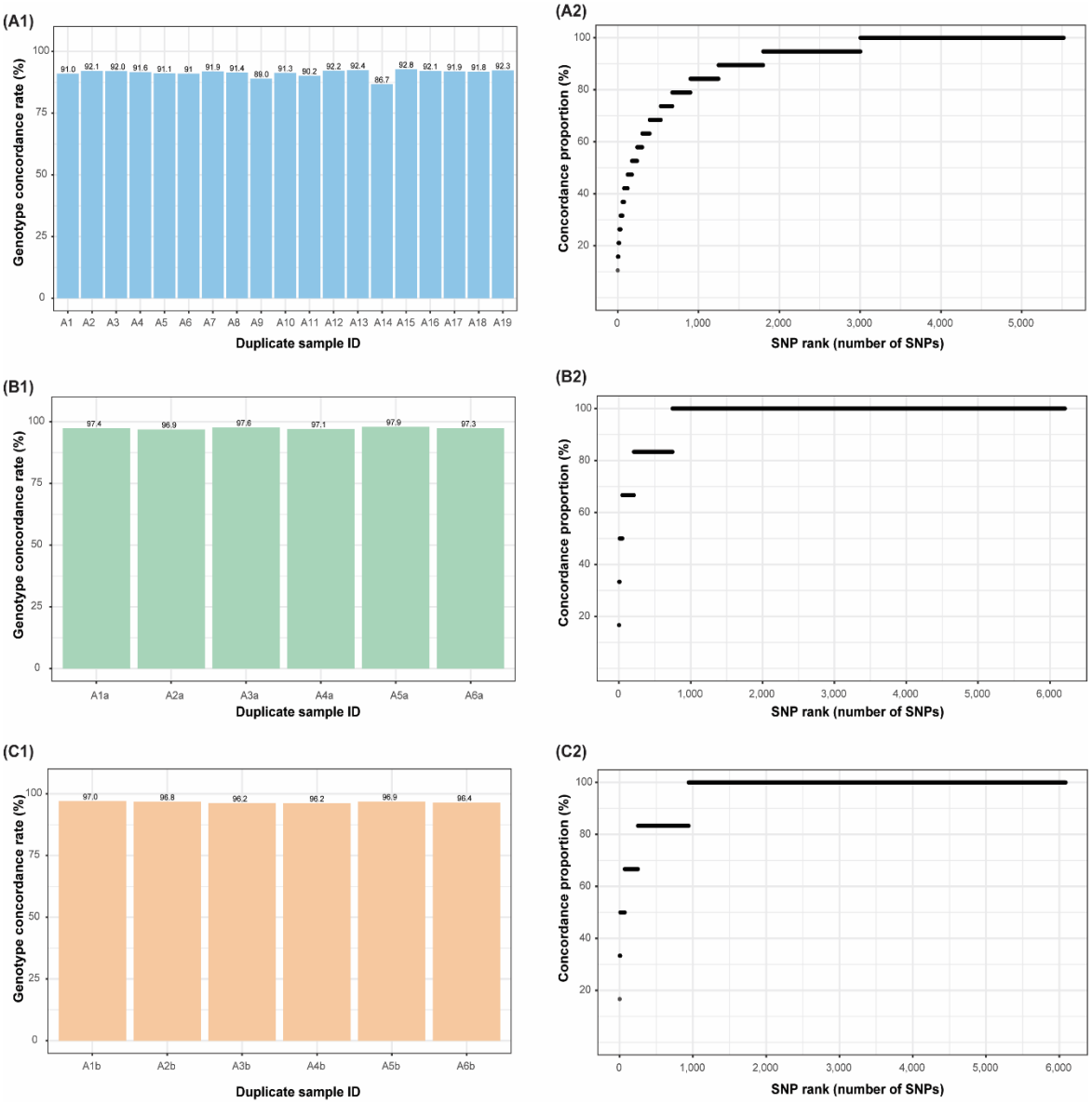


Figure 10 Genotype concordance across duplicate samples and SNP levels concordance for three types of comparisons. **(A1)** and **(A2)**: WGS vs. Allegro SNP panel (19 sample pairs, 5,519 SNPs). **(B1)** and **(B2)**: Within-batch concordance (6 sample pairs, 6,208 SNPs). **(C1)** and **(C2)**: Between-batch concordance (6 sample pairs, 6,085 SNPs). Bars represent the concordance rate per duplicate sample pair, and numbers above bars indicate exact percentage. SNPs were ranked by concordance proportion; each point represents one SNP.

Project objective 4:

Parentage assignment and pedigree reconstruction

Parentage analysis successfully assigned 615 of the 803 offspring (76.6%) to a candidate parent pair at 95% confidence, while 165 offspring (20.5%) were assigned to a single parent. Only 23 offspring (2.9%) could not be assigned to any candidate parent.

The assigned offspring were reconstructed into 191 full-sib families derived from 72 of the 87 candidate broodstock, demonstrating the effectiveness of the custom SNP panel for pedigree reconstruction. Broodstock reproductive contribution was highly variable, with individual parents contributing to 1–17 full-sib families and 1–204 assigned offspring.

Family representation was also highly uneven. More than half of the reconstructed families (106 of 191; 55.5%) were represented by a single offspring, whereas the largest family comprised 190 assigned offspring, indicating substantial variation in family contribution within the commercial breeding population.

Genetic parameters of growth and meat quality traits

Growth traits were measured in 803 Australian Akoya oysters, while meat quality traits were assessed in a subset of 414 oysters (**Table 2**). Growth traits exhibited moderate phenotypic variation, with fresh meat weight (FMW) and whole oyster weight (TOW) showing the greatest variation, whereas shell dimensions were comparatively uniform. Meat quality traits showed greater variation, particularly total glycogen content (TGC) and total lipid content (TLC), reflecting differences in both dry meat weight and tissue composition among oysters.

Heritability estimates indicated that all traits possess genetic variation and therefore have potential to respond to selection (**Table 2**). Growth traits showed low to moderate heritability (h^2 = 0.13 – 0.17), while glycogen concentration (GC) exhibited the highest heritability among the meat quality traits (h^2 = 0.49 ± 0.08). Log-transformed total glycogen content also showed moderate heritability (h^2 = 0.26 ± 0.08), whereas dry meat weight (DMW), lipid concentration (LC), and total lipid content (TLC) displayed relatively low heritability.

Table 2 Summary statistics and heritability estimates (h^2 ± SE) for growth and meat quality traits in Akoya oysters.

Traits	ID	n	Mean ± SD	Min	Max	CV (%)	h^2 ± SE
Whole oyster weight (g)	WOW	803	29.20 ± 9.44	10.50	62.50	32.33	0.16 ± 0.05
Fresh meat weight (g)	FMW	803	10.14 ± 3.86	3.00	24.40	38.04	0.13 ± 0.05
Anterior–posterior length (mm)	APL	803	58.84 ± 6.69	41.00	80.00	11.37	0.17 ± 0.05

Dorsal–ventral height (mm)	DVH	803	54.12 ± 6.41	28.00	76.00	11.84	0.15 ± 0.05
Shell depth (mm)	ShD	803	21.79 ± 2.74	15.00	31.00	12.59	0.17 ± 0.05
Dry meat weight (g)	DMW	414	1.54 ± 0.71	0.44	4.61	46.02	0.12 ± 0.07
Glycogen concentration (mg/g)	GC	414	88.02 ± 29.23	13.01	186.51	33.20	0.49 ± 0.08
Lipid concentration (mg/g)	LC	414	78.81 ± 17.09	47.62	162.79	21.69	0.18 ± 0.07
Total glycogen content (mg) †	TGC	414	142.94 ± 100.92	8.51	717.15	70.61	0.26 ± 0.08
Total lipid content (mg)	TLC	414	122.24 ± 64.55	25.92	378.74	52.80	0.10 ± 0.06

Descriptive statistics are presented on the original measurement scale. Concentrations are expressed as mg per g dry meat weight (DMW). † Total glycogen content was log-transformed prior to genetic analyses because of its positively skewed distribution. ID, traits abbreviation; n, number of samples; SD, standard deviation; CV, coefficient of variation; h^2, heritability; SE, standard error.

Genetic and phenotypic correlations are presented in **Figure 11**. Growth traits were strongly and positively genetically correlated, indicating that selection for the primary trait of increased oyster weight is expected to improve multiple growth-related traits simultaneously. In contrast, meat quality traits exhibited a mixture of positive and negative genetic correlations, suggesting more complex biological relationships. Notably, dry meat weight showed strong positive genetic correlations with both total glycogen and total lipid content, whereas glycogen and lipid concentrations were negatively correlated. These results indicate that multi-trait selection strategies may be required to achieve balanced improvement in oyster meat quality.

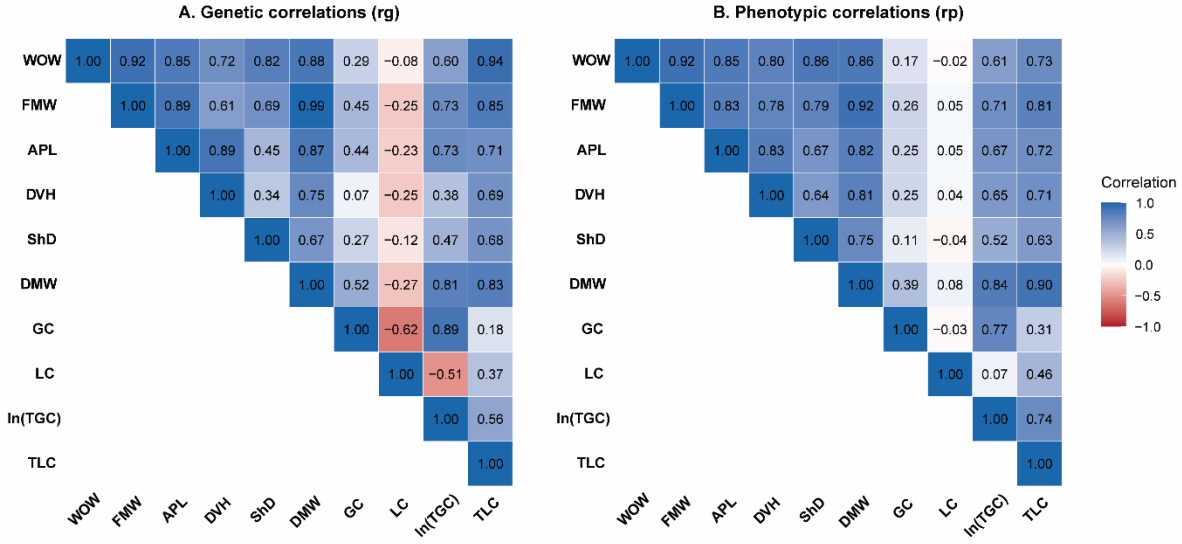


Figure 11 Genetic (**A**) and phenotypic (**B**) correlations among growth and meat quality traits. Correlations among growth traits were estimated using 803 oysters, whereas

correlations among meat quality traits and between growth and meat quality traits were estimated using a subset of 414 oysters. Trait abbreviations are defined in **Table 2**.

Genomic prediction of growth and meat quality traits

Genomic prediction performance for five growth and five meat quality traits was evaluated using six prediction methods (GBLUP, PBLUP, BayesC, mRKHS, RF, and XGBoost) based on 4,970 SNPs (**Figure 12**). Prediction accuracy ranged from 0.29 - 0.51 for growth traits and 0.43 - 0.74 for meat quality traits.

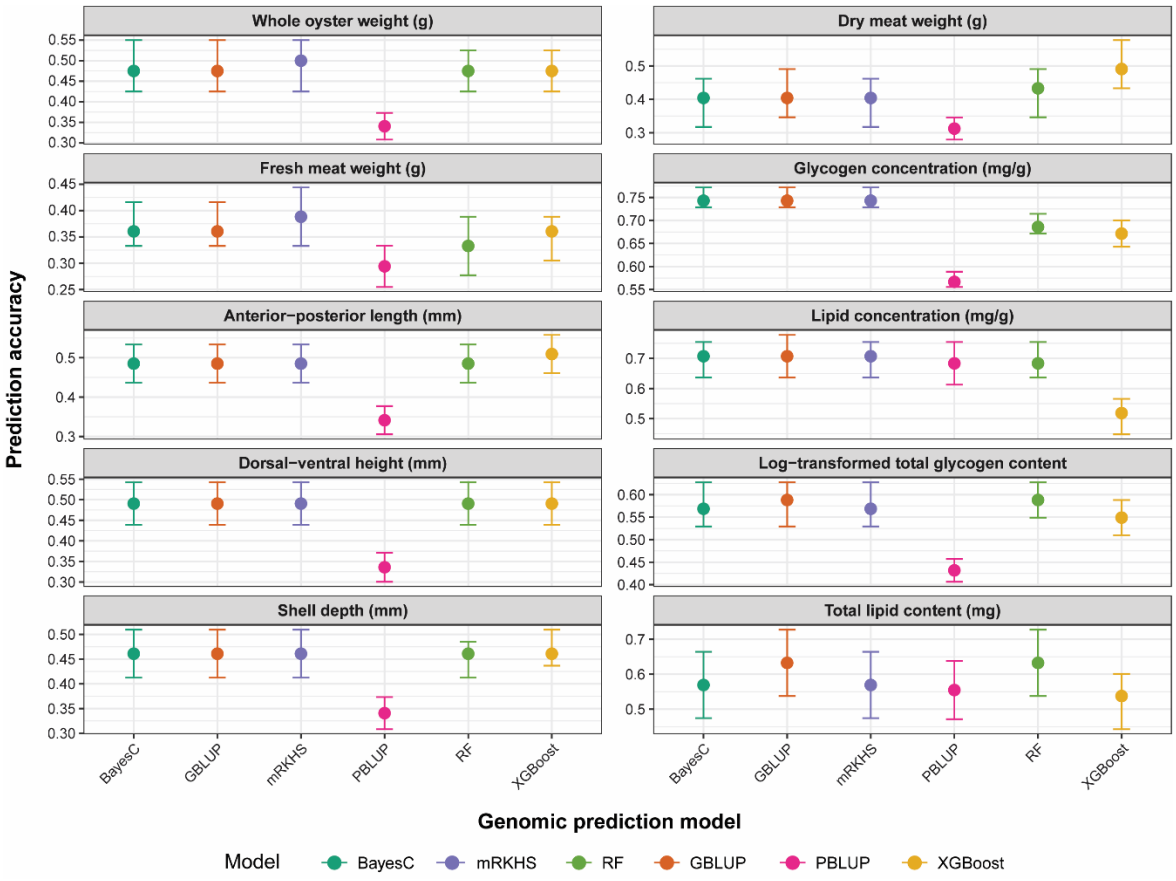


Figure 12 Prediction accuracy of genomic estimated breeding values (GEBVs) obtained using six prediction models, including genomic BLUP (GBLUP), pedigree-based BLUP (PBLUP), BayesC, multiple-kernel reproducing kernel Hilbert spaces regression (mRKHS), Random Forest (RF), and Extreme Gradient Boosting (XGBoost), for growth and meat-quality traits in Akoya oysters based on 4,970 SNPs. Prediction accuracy was estimated using fivefold cross-validation repeated 10 times (50 validation sets per trait). Each point represents the mean prediction accuracy across the 50 validation sets, and error bars indicate the 95% confidence intervals.

Among the growth traits, APL achieved the highest prediction accuracy across all models, while FMW showed the lowest prediction accuracy. For the meat quality traits, GC consistently achieved the highest prediction accuracy, followed by log-transformed TGC, whereas LC showed moderate prediction performance. Overall, traits with higher heritability generally achieved higher genomic prediction accuracy.

Prediction performance was broadly comparable among the six models, with substantial overlap in prediction accuracy across traits (**Figure 12**). GBLUP and mRKHS were the most consistently high-performing models, each achieving the highest mean prediction accuracy for six of the ten growth and meat-quality traits. BayesC, RF, and XGBoost also performed well, with each achieving the highest accuracy for four traits. Multiple models had the same highest accuracy for some traits. In contrast, PBLUP (traditional pedigree selection) did not achieve the highest accuracy for any trait and generally showed lower predictive performance than those based on genomic prediction. Overall, the genomic prediction models produced comparable results, and no single model performed best across all traits. These findings highlight the increased power and value of using genomic information and selecting the most suitable prediction model for each trait.

HIGHER DEGREE BY RESEARCH STUDENTS

This project supported the training of two research students at James Cook University: one BSc (Honours) student, who completed their degree during the project, and one PhD candidate, whose research and candidature is ongoing. Their research contributed directly to the development of innovative phenotyping technologies for Australian Akoya oyster selective breeding, including non-destructive near-infrared spectroscopy (NIR) and artificial intelligence (AI)-based image analysis. The outcomes will generate research publications and will continue to support the adoption of advanced phenotyping technologies in selective breeding programs.

Table 3. Higher Degree by Research students and their contributions to the project.

Students	Degree	Status	Contribution to the project	Outputs
Vinu Sankar	PhD	Ongoing	Development of automated image segmentation, automatic hinge alignment, AI-derived phenotypic	PhD in progress;

			descriptors (APL, DVH, shell and tissue descriptors), and machine learning models for predicting meat weight and whole oyster weight.	
Lenel Ahwan	BSc (Hons)	Completed (2026)	Development of NIR prediction models for glycogen and lipid content in Australian Akoya oysters	Thesis completed; manuscript in preparation for submission to a peer-reviewed journal.

CONCLUSIONS AND RECOMMENDATIONS

Conclusions

This project successfully developed an integrated suite of digital phenotyping and genomic resources to support selective breeding of Australian Akoya oysters. These resources include accurate AI-based models for shell dimensions and oyster weights, a validated targeted SNP array, genomic prediction models for growth and meat-quality traits, and integrated phenotype and genotype datasets. NIR spectroscopy also demonstrated potential for glycogen prediction in freeze-dried oyster tissue, although further development is required for fresh-tissue and lipid applications.

Collectively, the project has addressed several key technical barriers to industrial-scale selective breeding of Australian Akoya oysters by improving the efficiency of phenotypic data collection and providing cost-effective tools for parentage assignment, population genetic analysis, relatedness management, and genomic evaluation. Consequently, the project has delivered the enabling tools and foundational resources. Further research and industry investment are required to apply these outputs within a structured, multi-generation breeding program and demonstrate sustained genetic improvement under commercial conditions.

Recommendations

- **Characterise Australian Akoya oyster genetic resources:** A comprehensive population genomic assessment of Australian wild and cultured Akoya oyster populations is recommended. This work would guide the selection of genetically diverse founder broodstock, determine whether regional populations should be managed separately or combined, and support strategies to conserve diversity and minimise inbreeding.
- **Establish a selective breeding program:** Build on the AI phenotyping tools, SNP array, and genomic prediction models developed in this project to establish a structured selective breeding program with clearly defined industry breeding objectives.
- **Expand the reference population**: Continued collection of high-quality phenotype and genotype data across populations, environments, and generations is recommended. Expanding the reference population will improve the reliability of genomic prediction and enable evaluation of genetic gain, genotype-by-environment interactions, and prediction performance across commercial production settings.
- **Expand breeding objectives:** Extend breeding objectives beyond growth and meat quality to include economically important traits such as disease resistance, survival, reproductive performance, environmental resilience, and pearl quality.
- **Support industry adoption:** Further validate AI phenotyping and genomic technologies under commercial production conditions and continue developing NIR spectroscopy for routine meat-quality assessment. Establishing an integrated phenotype and genotype database will further support data-driven breeding decisions.

NEXT STEPS

The next phase of this research should focus on translating the tools and genomic resources developed through this project into a practical selective breeding program for Australian Akoya oysters. This will require continued collaboration between researchers and industry partners and further investment to validate the technologies at commercial scale.

The proposed next steps include:

- Develop a follow-on collaborative project with industry partners to establish the next phase of Australian Akoya oyster breeding.
- Undertake a nationwide population genomic survey of wild and cultured Australian Akoya oyster populations to identify genetically diverse founder broodstock and support long-term genetic management.

- Establish a reference breeding population and implement routine phenotyping, SNP genotyping, parentage assignment, and genomic evaluation using the tools developed in this project.
- Validate genomic selection under commercial farming conditions by monitoring genetic gain, diversity, and production performance over multiple breeding generations.
- Refine and integrate digital technologies, including AI phenotyping and NIR spectroscopy, into routine hatchery and farm operations to improve breeding efficiency and support industry adoption.

PROJECT TEAM

- Prof Dean Jerry, Project leader, James Cook University
- Prof Mostafa Rahimi Azghadi, James Cook University
- Dr Carla Ewels, James Cook University
- Dr Nga Vu, James Cook University
- Dr Cecile Massault, James Cook University
- Julie Goldsbury, James Cook University
- Anna Wilson, James Cook University
- Vinu Sankar, James Cook University
- Lenel Ahwan, James Cook University
- James Brown, Broken Bay Pearl farm
- Duncan Smith, Broken Bay Pearl farm
- Andrew Williamson, Broken Bay Pearl farm

APPENDICES

Appendices 1: Technical report: Development and evaluation of a custom Allegro SNP panel to support genetic improvement in the Australian Akoya oyster *Pinctada fucata*

Digital phenotyping to improve the efficiency of genomic selection in Akoya oysters

By Authors: Dean Jerry, Nga Vu, Carla Ewels, and Mostafa Rahimi Azghadi.
Date 06 August 2026

Project No. FA140

hello@foodagility.com
@FoodAgility
foodagility.com

REFERENCES

Berthelin, C., Kellner, K., & Mathieu, M. (2000). Storage metabolism in the Pacific oyster (Crassostrea gigas) in relation to summer mortalities and reproductive cycle (West Coast of France). *Comparative Biochemistry and Physiology Part B: Biochemistry and Molecular Biology*, *125*(3), 359-369.

Bligh, E. G., & Dyer, W. J. (1959). A rapid method of total lipid extraction and purification. *Canadian journal of biochemistry and physiology*, *37*(8), 911-917.

Brown, M. R., Kube, P. D., O'Connor, S., Cunningham, M., & King, H. (2012). Application of near-infrared reflectance spectroscopy for the rapid chemical analysis of Sydney rock oyster (Saccostrea glomerata) and Pacific oyster (Crassostrea gigas). *Journal of Shellfish Research*, *31*(4), 1051-1060.

Butler, D., Cullis, B., Gilmour, A., Gogel, B., & Thompson, R. a. (2017). ASReml-R reference manual version 4. *VSN International Ltd, Hemel Hempstead, HP1 1ES, UK*.

Carroll, N. V., Longley, R. W., & Roe, J. H. (1956). The determination of glycogen in liver and muscle by use of anthrone reagent. *J biol Chem*, *220*(2), 583-593.

Chen, D., Hu, B., Shao, X., & Su, Q. (2004). Removal of major interference sources in aqueous near-infrared spectroscopy techniques. *Analytical and bioanalytical chemistry*, *379*(1), 143-148.

Chen, T., & Guestrin, C. (2016). Xgboost: A scalable tree boosting system. Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining,

Chen, T., He, T., Benesty, M., Khotilovich, V., Tang, Y., Cho, H., Chen, K., Mitchell, R., Cano, I., Zhou, T., Li, M., Xie, J., Lin, M., Geng, Y., Li, Y., Yuan, J., & Cortes, D. (2026). *xgboost: Extreme Gradient Boosting*. In (Version 3.3.1.1) The R Foundation (CRAN). https://github.com/dmlc/xgboost

Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., Handsaker, R. E., Lunter, G., Marth, G. T., & Sherry, S. T. (2011). The variant call format and VCFtools. *Bioinformatics*, *27*(15), 2156-2158.

Du, X., Fan, G., Jiao, Y., Zhang, H., Guo, X., Huang, R., Zheng, Z., Bian, C., Deng, Y., & Wang, Q. (2017). The pearl oyster Pinctada fucata martensii genome and multi-omic analyses provide insights into biomineralization. *Gigascience*, *6*(8), gix059. https://doi.org/https://doi.org/10.1093/gigascience/gix059

Folch, J., Lees, M., & Stanley, G. S. (1957). A simple method for the isolation and purification of total lipides from animal tissues. *Journal of biological chemistry*, *226*(1), 497-509.

Fu, Q., Yu, Z., Zhao, J., Gao, L., Kong, N., Wang, L., & Song, L. (2024). Spatial compartmentalization and temporal stability of associated microbiota in Pacific oyster Crassostrea gigas. *Journal of Oceanology and Limnology*, *42*(4), 1348-1358.

Getachew, A. T., Lee, H. J., Cho, Y. J., Chae, S. J., & Chun, B. S. (2019). Optimization of polysaccharides extraction from Pacific oyster (Crassostrea gigas) using subcritical water: Structural characterization and biological activities. *International Journal of Biological Macromolecules*, *121*, 852-861.

Jones, O. R., & Wang, J. (2010). COLONY: a program for parentage and sibship inference from multilocus genotype data. *Molecular Ecology Resources*, *10*(3), 551-555. https://doi.org/https://doi.org/10.1111/j.1755-0998.2009.02787.x

Kalinowski, S. T., Taper, M. L., & Marshall, T. C. (2007). Revising how the computer program CERVUS accommodates genotyping error increases success in paternity assignment. *Molecular Ecology*, *16*(5), 1099-1106. https://doi.org/https://doi.org/10.1111/j.1365-294X.2007.03089.x

Li, H. (2011). A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics*, *27*(21), 2987-2993.

Li, Z., Qi, H., Yu, Y., Liu, C., Cong, R., Li, L., & Zhang, G. (2023). Near-infrared spectroscopy method for rapid proximate quantitative analysis of nutrient composition in Pacific oyster Crassostrea gigas. *Journal of Oceanology and Limnology*, *41*(1), 342-351.

Liaw, A., & Wiener, M. (2002). Classification and regression by randomForest. *R news*, *2*(3), 18-22.

Ma, X.-H., Chen, Z.-G., & Liu, J.-M. (2024). Wavelength selection method for near-infrared spectroscopy based on Max-Relevance Min-Redundancy. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*, *310*, 123933.

Min, Y., Dong, S., Su, M., Zhao, Y., & Zeng, M. (2020). Physicochemical, microbiological and sensory quality changes of tissues from Pacific oyster (Crassostrea gigas) during chilled storage. *Journal of Food Science and Technology*, *57*(7), 2452-2460.

Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., & Antiga, L. (2019). Pytorch: An imperative style, high-performance deep learning library. *Advances in neural information processing systems*, *32*.

Pérez, P., & de Los Campos, G. (2014). Genome-wide regression and prediction with the BGLR statistical package. *Genetics*, *198*(2), 483-495.

Prokhorenkova, L., Gusev, G., Vorobev, A., Dorogush, A. V., & Gulin, A. (2018). CatBoost: unbiased boosting with categorical features. *Advances in neural information processing systems*, *31*.

Team, O. (2025). Opencv: Open source computer vision library. *URL: https://opencv. org/. accessed*, 04-10.

Van Rossum, G., & Drake, F. L. (1995). *Python reference manual* (Vol. 111). Centrum voor Wiskunde en Informatica Amsterdam.

Vu, N. T., Jones, D. B., Massault, C., Zenger, K. R., Cooke, I., & Jerry, D. R. (2026). Tracking family contributions and family-level survival in silver-lipped pearl oyster *Pinctada maxima*: Unlocking cost-effective strategies for selective breeding. *Aquaculture*, *614*, 743560. https://doi.org/**https://doi.org/10.1016/j.aquaculture.2025.743560**

Wang, W., Yang, J., Li, Q., Ji, R., Gong, X., & Li, L. (2015). Development of calibration models for rapid determination of chemical composition of Pacific oyster (Crassostrea gigas) by near infrared reflectance spectroscopy. *Journal of Shellfish Research*, *34*(2), 303-309.

Wang, Z., Liu, Y., Zheng, Z., & Deng, Y. (2022). A genome-wide association study identifies SNPs and candidate genes associated with nacre color of pearl oyster *Pinctada fucata martensii*. *Aquaculture*, *561*, 738743. https://doi.org/https://doi.org/10.1016/j.aquaculture.2022.738743

Zheng, X., Levine, D., Shen, J., Gogarten, S. M., Laurie, C., & Weir, B. S. (2012). A high-performance computing toolset for relatedness and principal component analysis of

SNP data. *Bioinformatics*, *28*(24), 3326-3328. https://doi.org/https://doi.org/10.1093/bioinformatics/bts606

Zheng, Z. (2023). *The improved chromosome-level genome for the pearl oyster Pinctada fucata martensii using PacBio sequencing and Hi-C [Genome assembly ASM3311930v1; GenBank: GCA_033119305.1]* (https://www.ncbi.nlm.nih.gov/datasets/genome/GCA_033119305.1/

Zheng, Z., Hao, R., Yang, C., Jiao, Y., Wang, Q., Huang, R., Liao, Y., Jian, J., Ming, Y., & Yin, L. (2023). Genome-wide association study analysis to resolve the key regulatory mechanism of biomineralization in *Pinctada fucata martensii*. *Molecular Ecology Resources*, *23*(3), 680-693. https://doi.org/https://doi.org/10.1111/1755-0998.13743