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

Technical Report

By Nga Vu, Dave Jones, Ira Cooke, and Dean Jerry

Contents

List of Tables

List of Figures

Abbreviations

AGRF	Australian Genomics Research Facility
A_R	Allelic richness
BIC	Bayesian Information Criterion
CTAB	Cetyl-trimethyl ammonium bromide protocol
DAPC	Discriminant Analysis of Principal Components
ddRAD	double-digest RAD sequencing
F_{IS}	Inbreeding coefficient
F_{ST}	Population pairwise differentiation
GBS	Genotyping by Sequencing
gDNA	Genomic deoxyribonucleic acid
GRM	Genomic relationship matrix
GWAS	Genome-wide association study
H_E	Expected heterozygosity
H_O	Observed heterozygosity
HWE	Hardy–Weinberg equilibrium
IBD	Identical by decent
kbp	kilobase pairs (1,000 base pairs)
kNN	k-nearest-neighbour
LD	Linkage disequilibrium
MAF	Minor allele frequency
Mb	megabase (1,000,000 base pairs)
NCBI	National Center for Biotechnology Information
N_e	Effective population size
NGS	Next-generation sequencing
NSW	New South Wales
PC	Principal components
PPL	Percentage of polymorphic loci
QC	Quality control
QTLs	Quantitative trait loci
QUAL	Phred-scaled quality score
SD	Standard deviation
SNP	Single nucleotide polymorphisms
SPET	Single Primer Enrichment Technology
WA	Western Australia
WGS	Whole-genome sequencing

Acknowledgements

Development and evaluation of the custom Allegro SNP panel to support genetic improvement in the Australian Akoya oyster *Pinctada fucata* was conducted as part of the project funded by Food Agility CRC (FA140).

Akoya oyster samples were collected in 2022 under the Innovation Connection Grant (ICG002138), in 2023 under the Innovation Connection Grant (ICG002436), and during 2024-2025 as part of the Food Agility CRC (FA140) projects.

We would like to thank Julie Goldsbury and Hannah McCarthy for their assistance with laboratory work. We also thank the staff and students at James Cook University, including Sukkyoung Lee, Lenel Ahwan, Anna Wilson, and Carla Ewels, for their assistance with sampling. We are grateful to the staff of Broken Bay Pearl Farm (NSW), Harvest Road (WA), and the East 33 hatchery (NSW) for providing samples and support during sampling.

Executive summary

The Akoya oyster *Pinctada fucata* is an economically important aquaculture species, primarily valued for marine pearl production and increasingly recognised as a seafood resource. Despite growing interest in Australian Akoya aquaculture, genomic resources remain limited, particularly cost-effective genotyping tools suitable for routine breeding programs. Targeted genotyping-by-sequencing enables efficient interrogation of thousands of predefined SNP loci through selective enrichment of genomic regions, providing a rapid and cost-effective approach for SNP genotyping.

In this study, aquaculture Akoya oyster samples from New South Wales (NSW) and Western Australia (WA) were analysed using whole-genome sequencing (WGS) and double-digest RAD sequencing (ddRAD) to develop a custom low-density targeted SNP panel (Allegro SNP panel). Following multi-step filtering and probe design, 9,448 SNPs represented by 17,182 probes were incorporated, evenly distributed across the 14 chromosomes with an average spacing of ~109 kbp, ensuring broad genome coverage. The panel was validated on 190 Akoya oyster samples from four aquaculture sources, achieving 95.7% genotyping success and 97.9% polymorphic loci. Average sample and SNP call rates were high (90.4% and 90.7%), with ~97% internal repeatability and reproducibility and ~91% concordance with WGS data.

Genetic analyses using the panel successfully assigned both parents for most progeny, revealed moderate to high relatedness among broodstock, captured population-level diversity, and detected substantial differentiation between NSW and WA populations. These results demonstrate that the Allegro SNP panel exhibits high technical performance and is suitable for parentage assignment, population genetic analyses, and breeding applications in Australian Akoya oysters.

1. Introduction

The Akoya pearl oyster *Pinctada fucata* (also referred to as *P. fucata martensii* or *P. imbricata* or *P. radiata*) is widely distributed in China, Japan, Vietnam, and Australia and represents one of the most economically important species in marine pearl culture (O'Connor et al., 2003; Pit, 2004; Wada & Tëmkin, 2008). Commercial Akoya pearl culture was first established in Japan in 1916 and dominated the global market for several decades (Kuchel et al., 2011; Nagai, 2013). In the late 20th century, the industry suffered severe declines due to disease, dramatically reducing global pearl production by the early 2000s (Matsuyama et al., 2021; Miyazaki et al., 1999). In Australia, the potential for Akoya oyster aquaculture has been explored since the early 1990s, with experimental farming first established in New South Wales (NSW) and subsequent trials conducted in Queensland and Western Australia (WA), primarily for pearl production (Cropp et al., 2011; Kuchel et al., 2011; O'Connor et al., 2003; Pit, 2004). More recently, there has been increasing interest in Akoya oyster meat as an emerging seafood product due to its nutritional value and the presence of bioactive compounds (Theodorou et al., 2024; Wu et al., 2018). In addition, studies have shown that low-temperature pasteurization can maintain its physicochemical and microbiological quality, supporting safe consumption and commercialization (Chung et al., 2021).

Over the past decades, genetic tools for aquaculture and fishery management have advanced rapidly, driving significant improvements in breeding efficiency, stock management, and conservation efforts (Houston et al., 2020; Robledo et al., 2018). Among molecular markers, single nucleotide polymorphisms (SNPs) have become the most widely adopted in aquaculture research due to their high abundance across genomes, co-dominant inheritance, and cost-effective genotyping (Flanagan & Jones, 2019; Wenne, 2018; Zenger et al., 2019). The widespread use of SNP markers has been closely linked to advances in next-generation sequencing (NGS) technologies, which have enabled the rapid generation of large-scale genomic data in aquaculture species (Houston et al., 2020; Yang et al., 2025). In recent years, genotyping-by-sequencing (GBS) has emerged as a versatile high-throughput approach, enabling the rapid generation of genome-wide SNP data even in species with limited genomic resources (Houston et al., 2020; Robledo et al., 2018). While whole-genome sequencing (WGS) and GBS provide high-density genome-wide SNP data, they remain relatively costly for routine genotyping of large numbers of individuals in aquaculture breeding programs. To overcome these limitations, low- to medium-density SNP arrays have been developed, offering

a cost-effective, high-throughput, and scalable solution for targeted genotyping applications, such as parentage assignment, population genetic analyses, and genomic-informed selection (Guppy et al., 2020; Kriaridou et al., 2023; Lin et al., 2023; Massault et al., 2021; Qu et al., 2024). These targeted genotyping platforms leverage NGS technologies to capture and sequence a predefined set of informative SNPs, minimizing missing data and maximizing reproducibility across samples and loci. Among these, Single Primer Enrichment Technology (SPET) is a targeted sequencing approach in which a single primer anneals adjacent to each predefined SNP locus and initiates local extension during library preparation, enabling selective enrichment of genomic regions surrounding targeted variants (Scaglione et al., 2019; Leveque et al., 2025; Sung et al., 2024). In practice, SPET provides a cost-effective and flexible genotyping framework, supported by standardized, kit-based library preparation workflows that facilitate extensive sample multiplexing and the simultaneous analysis of large numbers of predefined SNP markers in a single sequencing run.

Despite growing interest in Akoya oyster aquaculture in Australia, genomic resources for Australian populations remain limited, particularly cost-effective genotyping tools suitable for routine breeding programs. By contrast, high-quality genetic resources have been developed internationally for *P. fucata*, including a haplotype-phased genome assembly and whole-genome resequencing (WGS) of wild and breeding populations, providing a foundation for genetic and breeding studies (Du et al., 2017; Takeuchi et al., 2022). In addition, 32 quantitative trait loci (QTLs) associated with seven growth traits and one sex trait were mapped across 10 linkage groups, with 695 candidate genes annotated (Liu et al., 2020). Genome-wide association studies (GWAS) in *P. fucata* have further dissected the genetic basis of biomineralization, shell and pearl growth, and nacre colour, identifying hundreds of significant SNPs and candidate genes with potential applications in selective breeding (Wang et al., 2022). Together, these studies highlight the increasing availability of genomic resources for Akoya oysters and their potential use in guiding selective breeding programs in Australia.

This report presents 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 is based on SPET. Using aquaculture Akoya samples from both NSW and WA, the study had two main objectives: (1) to develop a panel of informative SNPs suitable for selective breeding applications, and (2) to validate its performance in aquaculture populations. For SNP discovery, both WGS and ddRAD datasets were leveraged to identify high-quality and informative SNPs. The resulting custom Allegro

panel was then used to genotype 190 Akoya individuals, and its performance was evaluated through population genetic analyses as well as breeding and hatchery applications. By allowing the selection of a tailored set of SNPs and supporting extensive sample multiplexing in a single sequencing run, the Allegro platform provides a cost-effective and scalable genotyping solution suitable for aquaculture breeding programs. The development of this custom Allegro SNP panel represents a significant advancement, providing a cost-effective and powerful tool to support genomic selection in Australian Akoya oyster breeding programs.

2. Materials and methods

2.1 Akoya oyster sample collection and DNA extraction

Australian Akoya oyster samples used for the development of the custom SNP panel were collected from New South Wales (NSW) and Western Australia (WA) at three separate events (n = 234 individuals); Table 1). The first sample set was collected in 2022 from an oyster farm at Broken Bay, NSW (n = 99) under a previous project in 2022-2023. The second sample set was collected in 2023 from an oyster farm at Albany, WA (n = 115) under a previous project in 2023-2024. The third sample set was collected in 2024 comprised samples from the same farm at Broken Bay, NSW (n = 5 individuals), broodstock of wild origin from a hatchery at Camden Head, NSW (n = 7 individuals), and commercial product (frozen oysters) from the same farm at Albany, WA (n = 8 individuals).

To assess the genotyping performance of the custom Allegro panel developed in this study, a total of 190 samples representing 178 unique Akoya oysters from four sources, collected between 2024 and 2025, were genotyped in two batches using the custom Allegro panel (Table 1). These samples included 30 adult oysters from a farm at Broken Bay, NSW, 79 broodstock of wild origin from a hatchery at Camden Head, NSW, and 45 juvenile oysters (seven months old) from a farm at Broken Bay, NSW, and 24 adult oysters (commercial product) from a farm at Albany, WA. Twelve samples were intentionally duplicated for evaluating genotyping repeatability. Of these, six juvenile oysters were genotyped in duplicate within the same genotyping batch, while five juvenile oysters plus one broodstock individual were genotyped in duplicate across two different genotyping batches

The 45 juvenile oysters originated from a commercial mass spawning event conducted in October 2024 at a hatchery in Camden Head, NSW. A total of 87 broodstock oysters were used for the mass spawning, without prior identification of individual sex. Two weeks after the

spawning event, 75 broodstock were sampled using a non-invasive swabbing technique as described by Massault et al. (2022). The remaining 12 broodstock that may have contributed to the spawning event were not sampled due to either post-spawning mortality or failure to open during the swabbing procedure. Larvae were reared for 45 days until settlement on mesh bags and were subsequently transferred to longlines at a farm at Broken Bay, NSW. At seven months of age, 45 juvenile oysters were sampled at the farm, and whole-body tissue was collected for genotyping.

Akoya samples were collected for both the SNP panel development and validation using either a non-invasive swabbing technique as described by Massault et al. (2022), or collecting foot muscle tissue, depending on sample source (**Table 1**). Briefly, when swabbing, each adult Akoya oyster was swabbed four times on each side using a stiff comb with long, dense bristles (Picksters® size 7) to collect epithelial cells. The brush tips were then cut from the handles and placed into 2 mL screw-capped tubes containing RNAlater. Commercial frozen Akoya oyster products from a farm at Albany, WA were stored at –20 °C prior to shipment. Upon arrival at James Cook University (JCU), foot tissue was excised from frozen oysters and preserved in 100% ethanol, which was replaced with 70% ethanol after five days for long-term storage. Adult oysters collected from a farm at Broken Bay, NSW were sampled by excising foot tissue and preserving it in RNAlater. Broodstock samples were collected using non-invasive swabbing technique and preserved in RNAlater. Juvenile oysters were sampled as whole-body tissue, excluding the stomach and digestive tract, and preserved in 100% ethanol, followed by transfer to 70% ethanol for long-term storage. All samples were transported to JCU on ice and stored at –20 °C until DNA extraction.

Table 1 Summary of Akoya oyster samples and sequencing methods used for the custom Allegro SNP panel development and validation.

Use in study	Sequencing method	Population ID	Sample sources	Sampling location	Year of collection	No of samples	Sampling method
Custom Allegro SNP panel development	ddRAD sequencing	NSW2022_Comm	Commercial - Farm	Brooken Bay, NSW	2022	99	Non-invasive swabbing
		WA2023_Comm	Commercial - Farm	Albany, WA	2023	115	Non-invasive swabbing
	Whole genome sequencing	NSW2024_Comm	Commercial - Farm	Brooken Bay, NSW	2024	5	Foot muscle tissue
		NSW2024_Brds	Broodstock (wild)	Camden Head, NSW	2024	7	Non-invasive swabbing
		WA2024_Comm	Commercial product (frozen)	Albany, WA	2024	8	Foot muscle tissue
Custom Allegro SNP panel validation	Custom Allegro SNP panel	NSW2024_Comm	Commercial - Farm	Brooken Bay, NSW	2024	30	Foot muscle tissue
		NSW2024_Brds	Broodstock (wild)	Camden Head, NSW	2024	79	Non-invasive swabbing
		NSW2025_Juve	Juvenile oyster	Brooken Bay, NSW	2025	49	Foot muscle tissue
		WA2024_Comm	Commercial product (frozen)	Albany, WA	2024	24	Foot muscle tissue

** ddRAD: double-digest RAD sequencing; NSW: New South Wales; WA: Western Australia; NSW2022_Comm: commercial oysters from a NSW farm collected in 2022; WA2023_Comm: commercial oysters from a WA farm collected in 2023; NSW2024_Brds: broodstock of wild origin from NSW collected in 2024; NSW2025_Juve: juvenile oysters from a NSW farm collected in 2025.*

Genomic DNA (gDNA) from oyster samples collected during the previous research projects were extracted using the Chemagic™ DNA Extraction Kit LH (PerkinElmer Inc.) on a Zephyr™ automated liquid handling workstation (PerkinElmer Inc., Australia), following manufacturer's instructions (see Jerry et al. (2022)). gDNA from oyster samples used for whole genome sequencing (WGS) for SNP panel development was extracted using a modified cetyl-trimethyl ammonium bromide (CTAB) protocol (Adamkewicz & Harasewych, 1996; Vu et al., 2020). gDNA from frozen oyster samples was also extracted using the modified CTAB protocol. For SNP panel validation, gDNA from broodstock swab samples and juvenile oysters was extracted using the Chemagic™ Tissue 10 mg DNA Extraction Kit (Revvity) on a Chemagic™ 360 automated workstation (Revvity), following the manufacturer's instructions.

Extracted gDNA was quantified using a Quantus™ fluorometer and QuantiFluor® dsDNA Quantification kit (Promega Inc., Australia) followed by integrity confirmation using 0.8% agarose gel containing GelGreen (ThermoFisher Scientific Pty Ltd, Australia). gDNA samples were normalized to different concentration depending on the sequencing methods. For WGS, gDNA samples were normalized to ≥ 50 ng/µl (20 µl final volume). For double-digest RAD sequencing (ddRAD) sequencing, gDNA samples were normalized to ≥ 20 ng/µl (20 µl final volume). For custom Allegro SNP panel sequencing, gDNA samples were normalized to ≥ 20 ng/µl (30 µl final volume). All samples were then submitted to the Australian Genomics Research Facility (AGRF) for library preparation and sequencing.

2.2 SNP selection and custom Allegro SNP panel design

2.2.1 Library construction and sequencing for SNP discovery

ddRAD sequencing libraries for 214 Akoya oysters were prepared by AGRF using a ddRAD-based protocol adapted from Peterson et al. (2012). A total of 50 ng genomic DNA from each sample were digested with the restriction enzymes EcoRI and MspI and ligated with unique barcoded adapters compatible with the restriction site overhangs. The digested-ligated samples were purified, and size selection was performed using a BluePippin system (Sage Science) with fragment ranges of 280–342 bp or 280–375 bp. Libraries were then PCR-amplified using indexed primers. Sequencing was conducted on an Illumina NovaSeq 6000 platform with paired-end 150 bp reads, and BCL files were converted to FASTQ using the Illumina DRAGEN BCL Convert pipeline (v07.031.732.4.3.6). The raw sequencing data met AGRF quality standards and were used for downstream SNP calling and filtering.

WGS libraries for 20 individuals were constructed by AGRF using the Illumina DNA PCR-Free Prep Kit (Cat. No. 20041795) with 300 ng of genomic DNA per sample, following the manufacturer's protocol. Sequencing was performed on an Illumina NovaSeq X Plus System with paired-end 150 bp reads. Image analysis and real-time base calling were conducted using NovaSeq Control Software (NCS v1.2.2.48004) and Real Time Analysis (RTA v4.6.7). BCL files were subsequently converted to FASTQ using the Illumina DRAGEN BCL Convert pipeline (v07.031.732.4.3.6). The raw sequencing data met AGRF quality standards and were used for downstream SNP calling and filtering.

2.2.2 SNP calling and filtering

Variants calling for ddRAD and WGS dataset were carried out using a custom Nextflow workflow, Marine Omics Variant Pipeline (MOVP; https://github.com/marine-omics/movp/blob/main/README.md), which automates processing data from raw sequencing reads (fastq files) through to variant call format (vcf) files (Figure 1). Sequence alignment and SNP mapping were conducted using the *P. fucata martensii* genome. This project utilized the improved chromosome-level assembly published by Zheng (2023), available from the National Center for Biotechnology Information (NCBI) under accession number GCA_033119305.1 (ASM3311930v1). The original genome assembly was previously described by (Du et al., 2017). Briefly, paired-end reads of WGS and single-end reads of ddRAD were aligned to the *P. fucata martensii* reference genome (ASM3311930v1; GenBank: GCA_033119305.1) using BWA-MEM v0.7.17 (Li, 2013). To minimize biased estimates of allele frequencies, PCR duplicates were removed using MarkDuplicates (GTAK v4.6.1.0; McKenna et al. (2010)). Variant calling was performed within the MOVP framework. For ddRAD sequencing, variants were called using Freebayes v1.3.6 (Garrison & Marth, 2012) and SAMtools/mpileup v1.21 (Li et al., 2009). For WGS data, variant calling employed Freebayes v1.3.6 (Garrison & Marth, 2012), SAMtools/mpileup v1.21 (Li et al., 2009), and GATK HaplotypeCaller v4.6.1.0 (McKenna et al., 2010). The outputs from each variant calling workflow were initially filtered separately for each of the three sequencing datasets, to retained high-confidence SNPs before merging the outputs for each sequencing dataset (Figure 1).

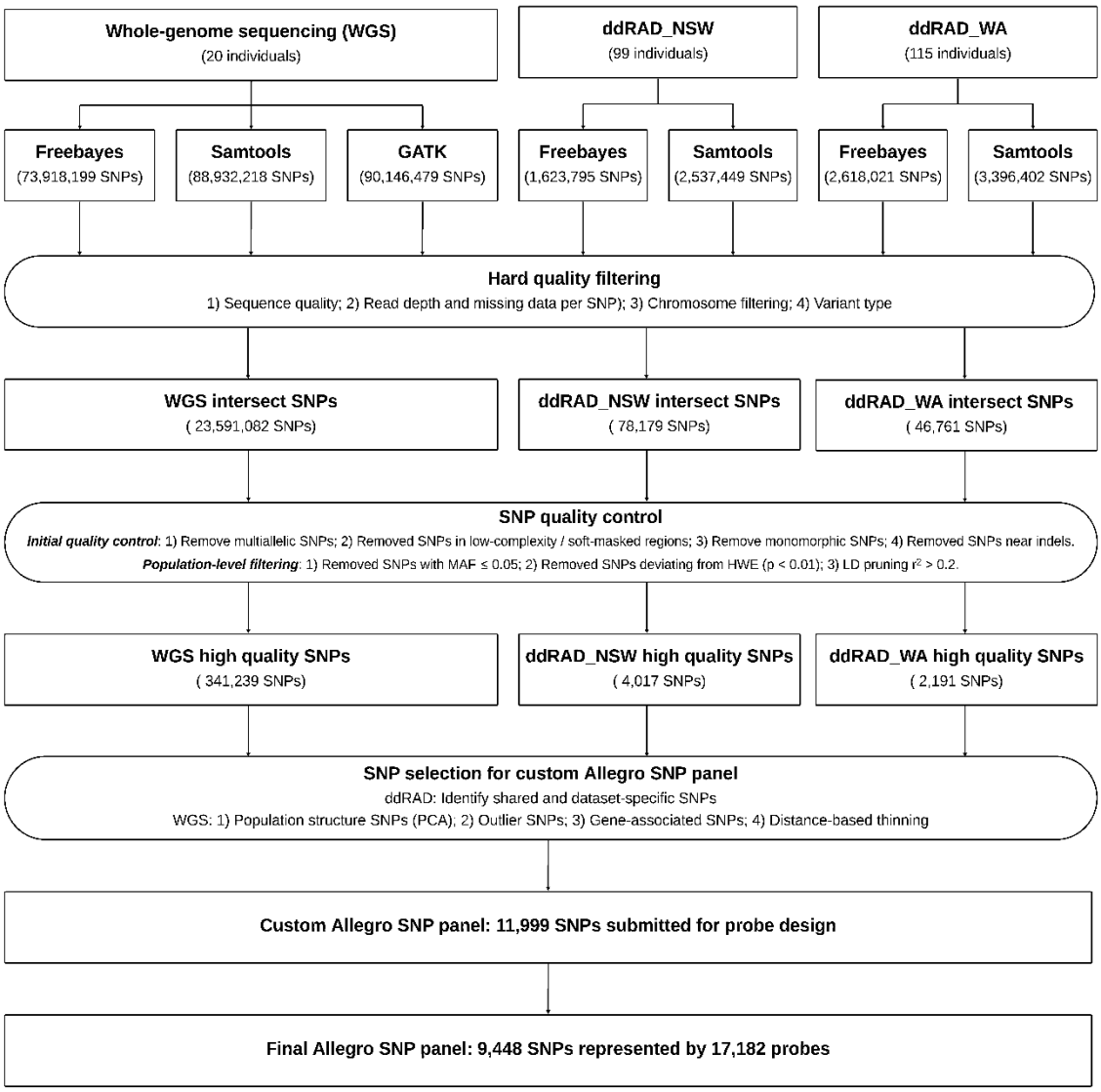

Figure 1 Workflow outlining the steps used for SNP calling, filtering, and selection of SNPs for the custom Allegro SNP panel.

The first filtering step involved hard filtering to remove low-quality variants based on the following criteria: (1) Phred-scaled quality score (QUAL) ≤ 20; (2) average sequence depth per sample < 5 or > 25 for the WGS dataset and average sequence depth per sample < 3 or > 25 for the ddRAD datasets; (3) missing data per variant ≥ 10%; (4) variants located outside the 14 assembled chromosomes (i.e., on unassigned contigs); (5) indels were removed to retained only SNPs. The remaining SNPs from this filter process of each sequence dataset were then merged to identify common SNPs for each sequence dataset using BCFtools v1.20 (Li, 2011). Additionally, samples with missing data > 20% were removed from each sequencing dataset. Consequently, the WGS dataset contained 23,591,082 SNPs across 19 individuals, the ddRAD-NSW dataset contained 78,179 SNPs across 99 individuals, and the ddRAD-WA dataset contained 46,761 SNPs across 109 individuals (Figure 1).

The second filtering step involved initial quality control and population-level filtering to identify high-quality SNPs for the target SNP panel. Initial quality control removed low-quality SNPs based on the following criteria: (1) multiallelic SNPs; (2) SNPs located in low-complexity regions & soft masked regions; (3) monomorphic SNPs; and (4) SNPs located within 10 bp of indels. Population-level filtering was then applied to remove SNPs with (1) minor allele frequency (MAF) ≤ 0.05, (2) significant deviation from Hardy–Weinberg equilibrium (HWE; $p < 0.01$), and (3) linkage disequilibrium (LD) pruning, retaining only one SNP from each pair with $r^2 > 0.2$.

Both the first and second filtering steps were conducted using BCFtools v1.20 (Li, 2011) and VCFtools v0.1.16 (Danecek et al., 2011). After these filtering steps, a total of 341,239, 4,017, and 2,191 SNPs were retained for the WGS, ddRAD-NSW, ddRAD-WA datasets, respectively, for SNP selection for the custom SNP panel (Figure 1).

2.2.3 SNP selection for the custom Allegro SNP panel

To select SNPs for the custom Allegro SNP panel, multiple selection steps were performed to identify informative SNPs suitable for application in Akoya oyster aquaculture selective breeding programs. This assessment included parentage assignment, genetic diversity assessment, population structure analysis, and genomic selection. To develop a custom Allegro SNP panel, our target was to submit a total of 12,000 high-quality candidate SNPs from both WGS and ddRAD sources for probe design.

First, SNPs associated with provenance analyses, intended for tracing individuals back to their sampling location, were identified according to the methods in Vu et al. (2023).The WGS dataset (n = 341,239 SNPs) was filtered to remove variants located within 30 kbp of each other using the "*--thin*" flag using VCFtools v0.1.16 (Danecek et al., 2011), thereby reducing LD caused by physical proximity. After this thinning step, 26,452 SNPs were retained for downstream analyses. For each SNP in this reduced dataset, we calculated pairwise genetic differentiation values (F_{ST}) using function *basic.stats* in the R package hierfstat v0.5.11 (Goudet, 2005). Two subsets of SNPs were then identified: (1) outlier SNPs detected using two independent population differentiation (PD) approaches: BayeScan v2.1 (Foll, 2012; Foll & Gaggiotti, 2008) and R package PCAdapt (Luu et al., 2017); (2) population-structure-informative SNPs, identified through Principal Component Analysis (PCA) using Discriminant Analysis of Principal Components (DAPC) in the R package adegenet v2.1.10

(Jombart, 2008; Jombart & Ahmed, 2012). From these analyses, SNPs associated with provenance were compiled by merging the identified lists and removing those with $F_{ST} < 0.05$.

Second, SNPs from ddRAD-NSW (*n* = 4,017 SNPs) and ddRAD-WA (*n* = 2,191 SNPs) datasets were assessed for inclusion as follows. These two datasets were merged to create a combined ddRAD dataset that included both shared and unique SNPs from each source. Variants located within 20 kbp of each other were then filtered out using the "*--thin*" flag in VCFtools v0.1.16 (Danecek et al., 2011). The combined ddRAD SNPs were then compared with the WGS dataset (*n* = 341,239 SNPs), and any SNPs do not present within the WGS dataset were removed.

Third, to identify SNPs associated with known genes, SNP position obtained from two published genome-wide association studies (GWAS; (Wang et al., 2022); Zheng et al. (2023)) were investigated. For each study, genomic coordinates of the reported SNPs were extracted, and a flanking region of ±50 bp around each SNP position was defined to account for positional uncertainty between reference genome versions. Two SNP lists were generated: one comprising 464 SNPs significantly associated with nacre colour (Wang et al., 2022) and another containing 252 SNPs associated with growth-related traits, including growth, total weight, and shell weight (Zheng et al., 2023). These SNP lists were then compared with the WGS dataset (*n* = 341,239 SNPs) to identify matching SNPs located within the defined ±50 bp windows. A total 200 SNPs identified through this process were considered candidate markers linked to previously reported genes associated with growth and nacre colour traits.

Finally, three SNP lists, including WGS SNPs associated with provenance analyses, ddRAD SNPs list, and SNPs associated with known genes, were merged to create a combined-SNP list that included both shared and unique SNPs from each source. A total of 4,516 SNPs were obtained after this merging process. Additional SNPs were then selected from the list of 26,452 WGS SNPs to fill genomic gaps between the 4,516 SNPs. These additional SNPs were chosen such that each was located at least 50 kbp away from any SNP in the 4,516-SNP list and at least 50 kbp apart from one another. This process yielded a final set of 11,999 SNPs, which was submitted to Tecan Genomics for probe design.

2.2.4 The custom Allegro SNP panel design

The custom Allegro SNP panel was developed by Tecan Genomics (Redwood City, United States), following the methodology outlined by Scaglione et al. (2019) and Andrews et al.

(2021). The chromosome-level assembly of *P. fucata martensii* (Zheng (2023); NCBI accession: GCA_033119305.1, ASM3311930v1) was used as reference genome. A list of 11,999 SNPs, including their genomic coordinates on the reference assembly, was submitted to Tecan Genomics in BED format for probe design. The Allegro panel employs Single Primer Enrichment Technology (SPET), which uses hybridization of custom-designed probes adjacent to each target SNP, followed by probe extension, adapter addition, and high-throughput sequencing on the Illumina platform (TECAN user guide for Allegro® Targeted Genotyping V2). For each targeted SNP, two probes were designed: one for hybridization to the plus strand and the other for the minus strand. These probes were positioned within approximately 100 bp of the target SNP to ensure efficient and accurate capture during library preparation.

2.3 Performance validation of custom Allegro SNP panel

2.3.1 Genotype performance and SNP quality

The genotyping performance of the custom Allegro SNP panel for Australian Akoya oysters was evaluated at both the sample and SNP levels using 190 individuals genotyped in two sequencing batches (Table 1). Sample-level performance was evaluated based on overall sample call rate, proportion of missing genotypes, and total sequencing reads per sample using BCFtools v1.20 (Li, 2011) and VCFtools v0.1.16 (Danecek et al., 2011). Samples with low call rates or high missingness were flagged as having poor DNA quality. Two samples failed due to low DNA quality and were removed prior to SNP-level performance assessment, leaving 188 samples.

SNP-level performance was evaluated using the remaining 188 samples. Metrics included SNP call rate, proportion of missing genotypes per SNP, and the number of polymorphic and monomorphic SNPs. Analyses were performed using VCFtools v0.1.16 (Danecek et al., 2011) and the R package vcfR v1.15.0 (Knaus & Grünwald, 2017). Genotypes were converted to numeric allele counts, and SNPs were classified as polymorphic if more than one unique genotype was observed across the 188 samples (excluding missing data), and monomorphic otherwise.

Following assessment of genotyping performance, quality control (QC) filtering was applied to ensure the reliability downstream analyses. Among the 188 samples, one replicate from each of the 12 duplicate pairs was removed prior to SNP filtering to avoid bias in allele frequency and diversity estimates, leaving 176 unique individuals. SNPs were then filtered

using BCFtools v1.20 (Li, 2011) and VCFtools v0.1.16 (Danecek et al., 2011) based on the following criteria: Phred-scaled quality score (QUAL ≥ 20), exclusion of indels and multiallelic loci, call rate ≥ 90%, MAF ≥ 0.02, and linkage disequilibrium (LD) pruning to one SNP from each pair exhibiting LD with $r^2 > 0.2$. The remaining 5,660 SNPs were considered high-quality and suitable for downstream analyses.

2.3.2 Evaluation of genotype accuracy, repeatability, and reproducibility

To evaluate the performance of the custom Allegro SNP panel, genotype concordance was assessed at both the sample and SNP levels to quantify accuracy, repeatability, and reproducibility. All concordance analyses were conducted starting from the raw dataset of 9,041 SNPs generated by the custom Allegro SNP panel. SNP filtering and concordance calculations were performed independently for each comparison dataset.

Genotype accuracy was evaluated by comparing SNP genotypes obtained from the custom Allegro SNP panel with WGS data. Nineteen individuals that were previously sequenced using WGS for SNP discovery and panel development were subsequently genotyped using the custom Allegro SNP panel. Genotype data from these individuals were merged, resulting in 38 genotype records (19 from WGS and 19 from Allegro panel). Prior to concordance analysis, SNP quality control was performed by removing indels and multi-allelic loci. To avoid bias due to missing genotypes, only SNPs with a call rate of 100% across all 38 genotype records were retained. After filtering, 5,519 SNPs were used for concordance analysis.

Genotype repeatability was assessed using six juvenile individuals that were genotyped twice within the same sequencing batch, resulting in 12 genotype records. SNP filtering was conducted as described above, retaining only biallelic SNPs with a call rate of 100% across tall 12 genotype records. After filtering, 6,208 SNPs were retained for concordance analysis.

Genotype reproducibility was evaluated using six individuals (five juveniles and one broodstock) that were genotyped independently in two separate sequencing batches, resulting in 12 genotype records (six per batch). SNP filtering was performed as described above, retaining only biallelic SNPs with a call rate of 100% across all 12 genotype records. After filtering, 6,085 SNPs were used for concordance analysis.

Sample-level concordance was calculated as the proportion of matching genotypes per individual pair using *bcftools gtcheck* in BCFtools v1.20 (Li, 2011). SNP-level concordance

was calculated by extracting genotypes from the filtered a VCF file and computing the proportion of identical genotypes calls (0, 1, or 2) across duplicate sample pairs. For each SNP, concordance was defined as the proportion of duplicate pairs with identical genotype calls, and SNPs were ranked based on this concordance. SNP concordance values were then visualized using the R package ggplot2 v4.0.2 (Wickham et al., 2016) and scales v1.4.0 (Wickham et al., 2025).

2.4 The application of custom Allegro panel in genetic improvement of Akoya oyster

2.4.1 Efficiency of the custom Allegro panel in parentage & relatedness analyses

To validate the power of the custom Allegro SNP panel for parentage assignment in Akoya oyster, broodstock and progeny from a commercial mass spawning event conducted in October 2024 at a hatchery in Camden Head, NSW, were genotyped (Table 1). A subset dataset comprising 118 individuals (73 broodstock and 45 progeny) and 5,660 SNPs was used for parentage assignment analyses (see Section 2.3.1). Although a large number of SNPs can improve resolution, using all 5,660 SNPs is computationally unnecessary for accurate pedigree reconstruction. Therefore, SNPs with a MAF < 0.3 were removed. Subsequently, SNPs located within 350 kbp of each other were filtered using the *--thin* function in VCFtools v0.1.16 (Danecek et al., 2011), resulting in a reduced dataset of 1,165 SNPs for parentage analyses.

For parentage assignment, broodstock relatedness were first estimated by estimating pairwise relatedness among all broodstock individuals through calculation of the genomic relationship matrix (GRM) using the R package SNPRelate v 1.42.0 (Zheng et al., 2012). GRM values were computed using 5,660 SNPs across 73 individuals. Based on expected theoretical relatedness coefficients and observed GRM distributions, values ≥ 0.40 were considered indicative of first-degree relationships (e.g., full-sibling), values between 0.15 and 0.40 were considered consistent with second-degree relationships (e.g., half-sibling), and values < 0.15 were interpreted as unrelated individuals.

Parent–progeny relationships among Akoya samples were subsequently inferred using CERVUS version 3.0.7 (Kalinowski et al., 2007) and COLONY version 2.0.7.0 (Jones & Wang, 2010) using 1,165 SNPs following the approach by (Vu et al., 2026). Briefly, CERVUS was run using default parameters, including a 1% genotyping error rate and 80% (relaxed) and 95% (strict) confidence levels. Parent pair analysis was conducted assuming unknown sex for candidate parents. COLONY analyses were performed assuming polygamy for both sexes, with

unknown parental sex, using a full-likelihood approach without sibship priors and a 1% genotyping error rate. Parentage assignments supported by both CERVUS and COLONY at the 95% confidence level were considered valid. In cases where assignments were identified by only one program but achieved 95% confidence, additional validation was performed by examining identical-by-descent (IBD) coefficients estimated using PLINK version 1.9 (Purcell et al., 2007) and GRM.

2.4.2 Population genetic diversity analyses

Genetic diversity was assessed in aquaculture Akoya oyster samples from NSW and WA using 5,566 SNPs (see Section 2.3.1). The dataset comprised 176 individuals (Table 1), including NSW2024_Comm (n = 30), NSW2024_Brds (n = 77), NSW2025_Juve (n = 45), and WA2024_Comm (n = 24). Population-level diversity metrics, including allelic richness (A_R), expected heterozygosity (H_E), and observed heterozygosity (H_O) were calculated using the *divBasic* function of the R package diveRsity v1.9.90 (Keenan et al., 2013), with 10,000 bootstrap replicates. Percentage of polymorphic loci (PPL) and inbreeding coefficient (F_{IS}) were computed using the gl.report.*heterozygosity* function of the R package dartR v2.9.9.5 (Mijangos et al., 2022). Effective population size (N_e) representing the effective number of breeders contributing to each sampled cohort, was estimated for each population using NeEstimator v2.1 based on the linkage disequilibrium (LD) method under a random mating model (Do et al., 2014).

2.4.3 Population genetic differentiation and structure analyses

Population differentiation and genetic structure were calculated for three aquaculture Akoya populations, including NSW2024_Comm (n = 30), NSW2024_Brds (n = 77), and WA2024_Comm (n = 24) using 5,660 SNPs. Genetic differences (F_{ST}) among populations were estimated using the *gl.fst.pop* function of the R package dartR v2.9.9.5 (Mijangos et al., 2022), with 1000 bootstrap replicates. Pairwise F_{ST} values and their significance between populations, were calculated according to Wright (1949) and Weir and Cockerham (1984).

Population genetic structure was determined using two approaches. First, Discriminant Analysis of Principal Components (DAPC) was performed using the R package adegenet package v2.1.11 (Jombart, 2008) to describe genetic differentiation among pre-defined groups based on individual SNP genotypes. The optimal number of genetic clusters was first explored

using the *find.clusters* function, which applies sequential k-means clustering with increasing numbers of clusters (K) and evaluates alternative solutions using the Bayesian Information Criterion (BIC). The number of principal components (PCs) retained for the DAPC was determined using the *optim.a.score* function, which identifies the optimal number of PCs that maximizes the discrimination power of the analysis while minimizing overfitting. Second, individual-level genetic relationships were inferred without prior population assignments using the R package NetView v1.0 (Neuditschko et al., 2012; Steinig et al., 2016). Using SNP-based genetic distances, NetView was applied across k-nearest-neighbour (kNN) values ranging from 1 to 60 to examine genetic structure from fine- to broad-scale resolution.

3. Results

3.1 SNP calling, filtering, and selection for the custom SNP panel design

WGS was performed for 20 Akoya individuals, generating a total of 904,477,772 paired-end reads, with an average of 45,223,889 read pairs per sample. In addition, ddRAD sequencing of 214 Akoya samples produced 2,802,188,357 single-end reads, with an average of 13,094,338 reads per sample. Raw sequence reads were aligned to the Akoya reference genome (Du et al., 2017; Zheng, 2023), with mapping rates ranging from 65.2% to 98.4% for the WGS dataset and 84.4% to 97.1% for the ddRAD dataset.

For the WGS dataset, variant calling using FreeBayes, SAMtools, and GATK HaplotypeCaller identified 73,918,199, 88,932,218, and 90,146,479 variants, respectively (Figure 1; Table S1). For the ddRAD-NSW and ddRAD-WA datasets, variant calling using FreeBayes identified 1,623,795 and 2,618,021 variants, respectively, while SAMtools identified 2,537,449 and 3,396,402 variants (Figure 1; Table S1). SNP quality control was conducted for each of the three sequencing datasets separately.

Following initial quality control and merging variant calls from FreeBayes, SAMtools, and GATK, the WGS dataset contained 23,591,082 SNPs across 19 individuals, after removing one individual with more than 20% missing genotypes (Table S1). Subsequent filtering retained 341,239 high-quality SNPs. After removing SNPs located within 30 kbp of each other, 26,452 SNPs remained. From this filtered dataset, two complementary approaches were used to identify SNPs associated with provenance differentiation. First, 757 outlier SNPs were detected using two independent population differentiation approaches (BayeScan and PCAdapt). Second, 994 unique SNPs were identified by combining the first six Principal Components

(PC1-PC6) in DAPC analysis. Together, these two SNP sets (757 outlier and 994 unique SNPs) were retained as candidate SNPs for custom SNP panel design.

For the ddRAD-NSW dataset, initial quality control and merging of variant calls from FreeBayes and SAMtools resulted in 78,179 SNPs across 99 individuals, which were further filtered to retain 4017 high-quality SNPs suitable for SNP panel development (Table S1). Similarly, the ddRAD-WA dataset contained 46,761 SNPs across 109 individuals after removal of six individuals with more than 20% missing genotypes and was subsequently filtered to retain 2,191 SNPs (Table S1). After merging the two ddRAD datasets, removing duplicated SNPs and SNPs located within 20 kbp of each other, 3858 SNPs remained. Comparison with the WGS dataset revealed that 1,063 SNPs were absent from the WGS data and were therefore excluded, resulting in a final set of 2,795 ddRAD-derived SNPs retained as candidate SNPs for custom SNP panel design.

In total, 757 outlier SNPs and 994 unique SNPs identified from the WGS dataset, together with 2,795 SNPs derived from ddRAD sequencing, were retained as candidate markers for the custom SNP panel design. In addition, 200 functional SNPs previously identified from GWAS studies and reported to be associated with growth and nacre colour traits were included (Wang et al., 2022; Zheng et al., 2023). These candidate SNPs were pooled to generate a unified SNP set comprising 4,516 SNPs. To reach the target of 11,999 SNPs for probe design, an additional 7,483 SNPs were selected from the WGS dataset (n = 26,452 SNPs), each located at least 50 kbp away from any SNP in the 4,516-SNP set and at least 50 kbp apart from one another. The 11,999 SNPs selected for probe design were relatively evenly distributed across all 14 chromosomes (Figure S1).

3.2 The custom Allegro SNP panel design

The target SNP panel was manufactured by Tecan Genomics (Redwood City, United States). Of the 11,999 SNPs submitted for panel design, 9,448 SNPs were successfully incorporated into the final panel, corresponding to an overall design coverage of 78.7%. Among these, 7,734 SNPs were represented by two independent probes, while 1,714 SNPs were represented by a single probe, resulting in a total of 17,182 probes processed for downstream genotyping. The number of SNPs per chromosome ranged from 356 to 1230, and the average distance between adjacent SNPs varied from 86,515 bp to 147,075 bp, with a genome-wide average of 108,937 bp, indicating a relatively even distribution of SNPs across chromosomes (Figure 2).

Of the 9,448 SNPs retained in the final panel, 1,435 originated from the ddRAD-NSW dataset, 2,037 from the ddRAD-WA dataset, and 132 were functional SNPs previously associated with growth and nacre colour traits. The inclusion of ddRAD-derived SNPs enhances genome-wide representation across populations, while the functional SNPs provide targeted markers relevant to selective breeding applications.

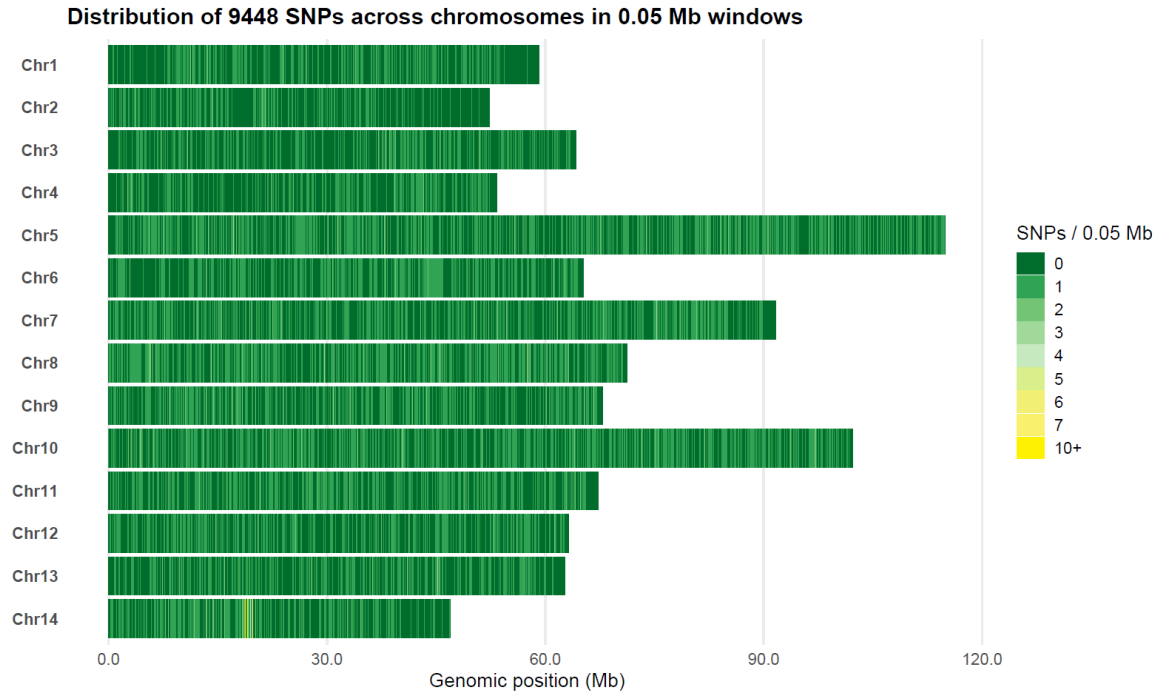

Figure 2 Genome-wide distribution of the 9,448 SNPs included in the custom Allegro SNP panel across 14 chromosomes. SNP positions were divided into non-overlapping 0.05 Mb windows. Tile colour indicates the number of SNPs per window, with counts capped at 10.

3.3 Performance validation of the custom Allegro SNP panel

3.3.1 Genotype performance and SNP quality

Of the 9,448 SNPs included in the final custom Allegro SNP panel design, 9,041 SNPs (95.7%) were successfully genotyped across 190 individuals processed in two genotyping batches. Total read depth per sample ranged from 45,558 to 8,244,075 reads, with an average of 1,618,258 ± 1,275,394 (mean ± SD; **Figure 3A**). Across the 190 individuals and 9,041 SNPs, sample call rate ranged from 49.7 to 94.8% with an average of 90.4 ± 3.8% (**Figure 3B**), while missing data per sample ranged from 5.2% to 50.3%, with an average of 9.6 ± 3.8%. Two individuals exhibited low call rates (64.9% and 49.7%), likely due to poor DNA quality, and were removed prior to SNP-level performance assessment, leaving 188 individuals for further evaluation.

Across 9,041 SNPs and the 188 individuals, SNP call rate ranged from 0.5 to 99.5% with an average of 90.7 ± 21 % (**Figure 3C**), while missing data per SNP ranged from 0% to 100%, with an average of 9.3 ± 21%. Among the 9,041 SNPs, 296 indels (3.3%), 926 multiallelic SNPs (10.2%), 185 monomorphic SNPs (2.1%), and 8,856 polymorphic SNPs (97.9%) were identified prior to quality filtering. SNP QC filtering was subsequently performed on the dataset of 9,041 SNPs across 176 unique individuals (after removal of duplicate samples; see Section 2.3.1; **Table S2**). Removal of indels and multiallelic loci retained 7,819 SNPs. A further 1,379 SNPs were removed due to missing data per SNP > 10% (call rate < 90%), 248 SNPs were excluded due to MAF < 0.02, and 478 SNPs were removed during LD pruning ($r^2 > 0.2$). This process retained 5,660 high-quality SNPs (62.6% of the raw 9,041 SNPs) across 176 individuals for downstream analyses (**Table S2**). The MAF distribution of the 5,660 high-quality SNPs retained after QC filtering of the custom Allegro SNP panel (**Figure 3D**).

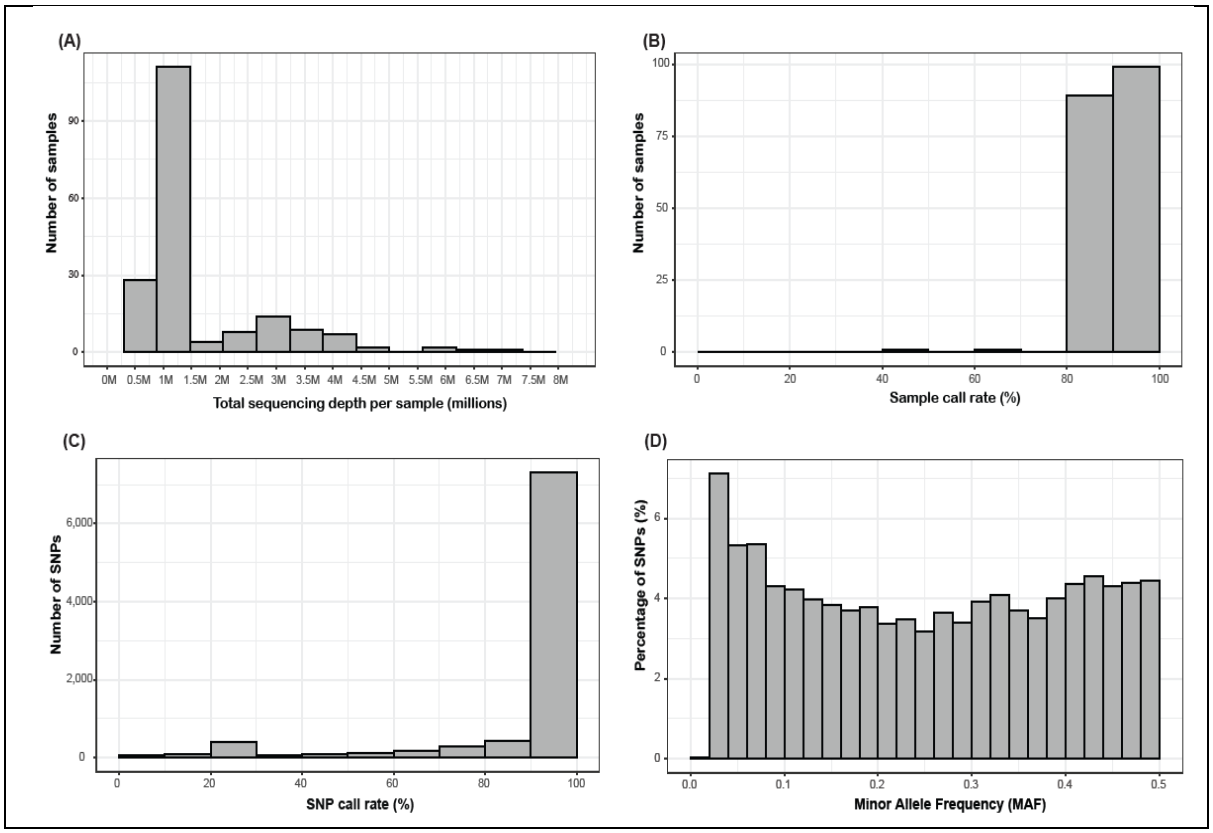

Figure 3 Genotyping performance and quality control of the custom Allegro SNP panel. **(A)** Distribution of total sequencing depth per sample across 190 individuals. **(B)** Distribution of sample call rate for the 9,041 raw SNPs across 188 individuals. **(C)** Distribution of SNP call rate for the 9,041 raw SNPs across 188 individuals after removing two low-quality samples. **(D)** MAF distribution for the high-quality 5,660 SNPs retained after quality control filtering of the custom Allegro SNP panel.

3.3.2 Genotype accuracy, repeatability, and reproducibility

Genotype concordance of the custom Allegro SNP panel was high at both the sample and SNP levels across three types of comparisons: accuracy, repeatability, and reproducibility (**Figure 4**; **Table S3**).

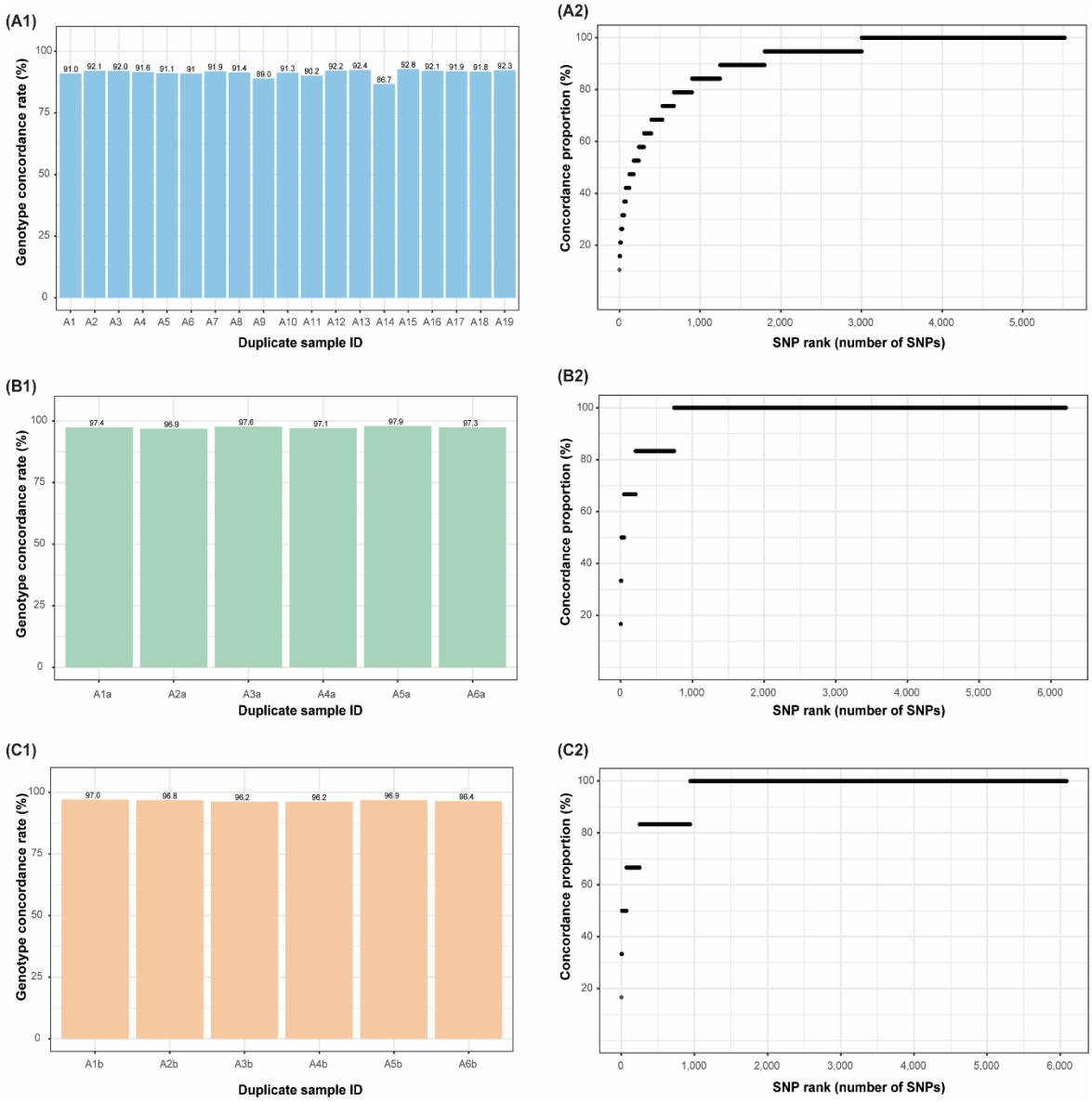

Figure 4 Genotype concordance across duplicate samples and SNP levels concordance for three types of comparisons. **(A1)** and **(A2)**: WGS vs. Allegro SNP panel (19 sample pairs, 5,519 SNPs). **(B1)** and **(B2)**: Within-batch concordance (6 sample pairs, 6,208 SNPs). **(C1)** and **(C2)**: Between-batch 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.

Genotype accuracy was assessed across 19 sample pairs using both WGS and the custom Allegro SNP panel, based on 5,519 SNPs with complete genotype calls (100% call rate). Sample-level genotype concordance ranged from 86.7% to 92.8%, with a mean of 91.3 ± 1.4% (**Figure 4A1**). At the SNP-level, concordance across the 19 sample pairs ranged from 10.5% to 100%, with a mean of 90.9 ± 14.3% (**Figure 4A2**). The majority of samples and SNPs were highly concordant between the two platforms, indicating strong genotype accuracy of the custom Allegro SNP panel relative to WGS.

Genotype repeatability was evaluated across six sample pairs genotyped twice within the same sequencing batch, using 6,208 SNPs with complete genotype calls. Sample-level genotype concordance ranged from 96.9% to 97.9%, with a mean of 97.4 ± 0.3% (**Figure 4B1**). At the SNP-level, concordance ranged from 16.7% to 100%, with a mean of 97.3 ± 8.3% (**Figure 4B2**).

Genotype reproducibility was assessed across six sample pairs genotyped independently in two separate sequencing batches, using 6,085 SNPs with complete genotype calls. Sample-level genotype concordance ranged from 96.2% to 97.0%, with a mean of 96.6 ± 0.3% (**Figure 4C1**). At the SNP-level, concordance ranged from 16.7% to 100%, with a mean of 96.2 ± 9.2% (**Figure 4C2**). The high concordance observed in both repeatability and reproducibility comparisons confirms the accuracy, reliability, and overall robustness of the custom Allegro SNP panel for genotyping.

3.4 Genetic analyses using the custom Allegro panel

3.4.1 Parentage assignment and family genetic structure

Pairwise genomic relatedness among the 73 broodstock individuals was estimated using GRM. Across 2,628 pairwise comparisons, GRM values ranged from −0.24 to 0.81, with a mean of −0.01 and a standard deviation of 0.17 (**Figure S2**). While most pairwise comparisons showed low relatedness, elevated relatedness was detected among several broodstock pairs. In total, 28 broodstock individuals were involved in at least one pairwise comparison with GRM > 0.40, consistent with potential full-sib relationships. A further 40 individuals showed moderate relatedness, with GRM values ranging from 0.20 to 0.38 in at least one pairwise comparison, indicating that a substantial proportion of the broodstock shared some degree of genetic relatedness.

Parentage assignment was successful for 36 of the 45 Akoya offspring (80%), which were assigned to full-sib parent pairs with 95% confidence. The remaining nine progeny were assigned to a single candidate parent with 95% confidence. The 36 assigned offspring were distributed across 33 full-sib families and involved a total of 37 broodstock as candidate parents. Because the sex of broodstock was unknown, parent pairs could not be definitively classified as maternal or paternal. Instead, CERVUS and COLONY identified parents as first and second candidate parents. Among the contributing broodstock, 15 individuals were identified as first candidate parents and 24 as second candidate parents, with two individuals appearing in both categories. The inferred family structure of the 33 full-sib families is presented in **Figure 5**.

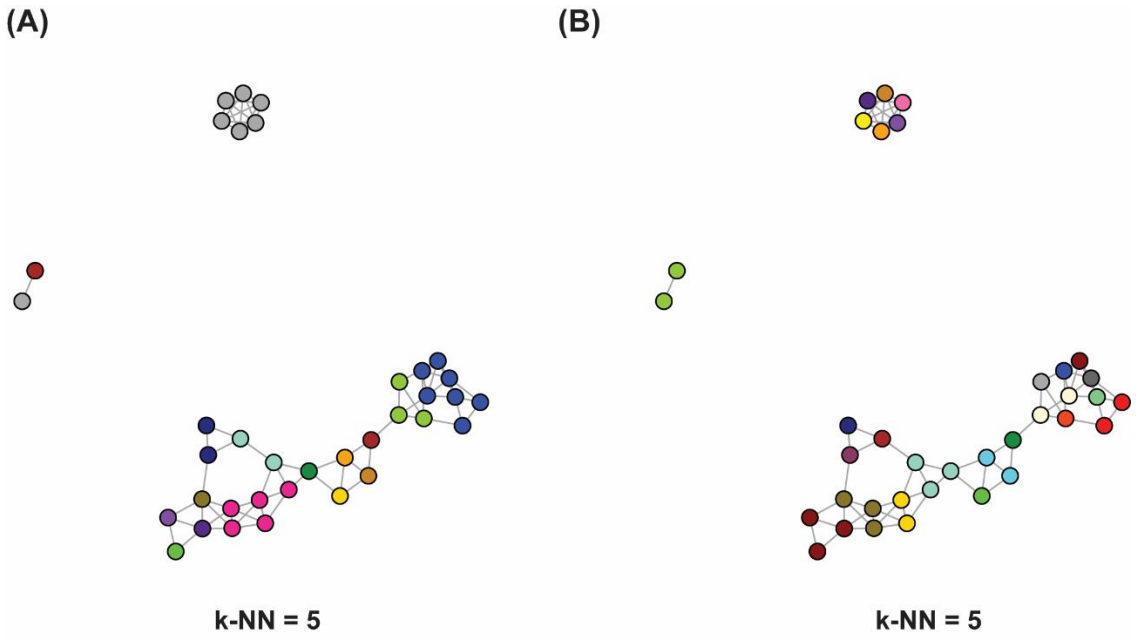

Figure 5 Family networks of the Akoya oyster *Pinctada fucata* progeny (n = 36) visualized using 5,660 SNPs and the Netview R package. The network analysis plot shows the family structure of the progeny as assigned to first candidate parent **(A)** and second candidate parent **(B)**. The maximum number of nearest neighbour (k-NN) threshold presented the number of broad genetic clusters.

3.4.2 Population genetic diversity

Using 5,660 SNPs from the custom Allegro panel, the results indicated that genetic diversity estimates of the three NSW Akoya aquaculture populations were broadly similar, but and slightly higher than that of WA2024_Comm population (**Table 2**).

Average PPL per population ranged from 95.67 to 99.10% across the NSW populations. In contrast, the WA population, WA2024_Comm showed a slightly lower value (88.73%). Average A_R per population ranged from 1.96 to 1.99 across the NSW populations, whereas the WA2024_Comm showed a lower value of 1.89. Average H_E and H_O were also similar among the NSW populations, differing by approximately 0.01 (H_E = 0.31 - 0.32; H_O = 0.30 - 0.31). In the WA2024_Comm population, H_E and H_O were slightly lower, at 0.30 and 0.29, respectively. Low inbreeding coefficients were observed across all four populations, with F_{IS} values ranging from 0.04 to 0.07.

Estimates of N_e based on linkage disequilibrium were relatively small for the NSW2024_Comm and NSW2024_Brds populations (Ne ≈ 9; 95% CI: ~7–12). In contrast, larger N_e estimates were observed for the NSW2025_Juve population (N_e = 40.6; 95% CI: 33.9–49.7) and the WA2024_Comm population (N_e = 46.9; 95% CI: 29.9–115.0).

Table 2 Genetic diversity of four Akoya oyster *P. fucata* aquaculture populations sampled

Populations	n	PLL	A_R	H_E	H_O	F_{IS} ($p < 0.05$)	N_e linkage disequilibrium	
							N_e	95% CI (lower-uper)
NSW2024_Comm	30	95.67	1.96 ± 0.20	0.31 ± 0.17	0.31 ± 0.21	0.04	8.8	6.7 - 11.6
NSW2024_Brds	77	99.10	1.99 ± 0.09	0.31 ± 0.16	0.31 ± 0.19	0.06	9	6.9 - 11.4
NSW2025_Juve	45	96.31	1.96 ± 0.19	0.32 ± 0.16	0.30 ± 0.18	0.05	40.6	33.8 - 49.7
WA2024_Comm	24	88.73	1.89 ± 0.32	0.30 ± 0.17	0.29 ± 0.19	0.07	49.6	29.9 - 115.0

n, number of samples; PPL, percentage of polymorphic loci; A_R, mean allelic richness; H_E, average expected heterozygosity; H_O, average observed heterozygosity; F_{IS}, inbreeding coefficient (significant at $p < 0.05$); N_e, effective population size based on linking disequilibrium SD, standard deviation.

3.4.3 Population differentiation and genetic structure

Population pairwise differentiation (F_{ST}) among the three Akoya aquaculture populations sampled ranged from 0.001 to 0.17, and all pairwise comparisons were significant ($p < 0.05$, **Table S3**). Pairwise F_{ST} between the two NSW populations and the WA population were 0.15 and 0.17, respectively, indicating moderate differentiation. In contrast, pairwise F_{ST} between the two NSW populations was very low (F_{ST} = 0.001, suggesting they are essentially genetically identical, despite the comparison being statistically significant (p = 0.017).

Both NetView and DAPC analyses indicated that aquaculture Akoya oysters from NSW and WA formed two main genetic clusters (**Figure 6**), with all NSW individuals clustering together and WA individuals forming a separate group. K-means clustering analysis implemented in the R package adegenet identified five genetic clusters (K = 5) as the optimal number based on the lowest BIC value and the *optim.a.score* function. Within the NSW populations, individuals from NSW2024_Brds and NSW2024_Comm were further divided into four subclusters, whereas all individuals from the WA2024_Comm population formed a single cluster (**Figure 6A**).

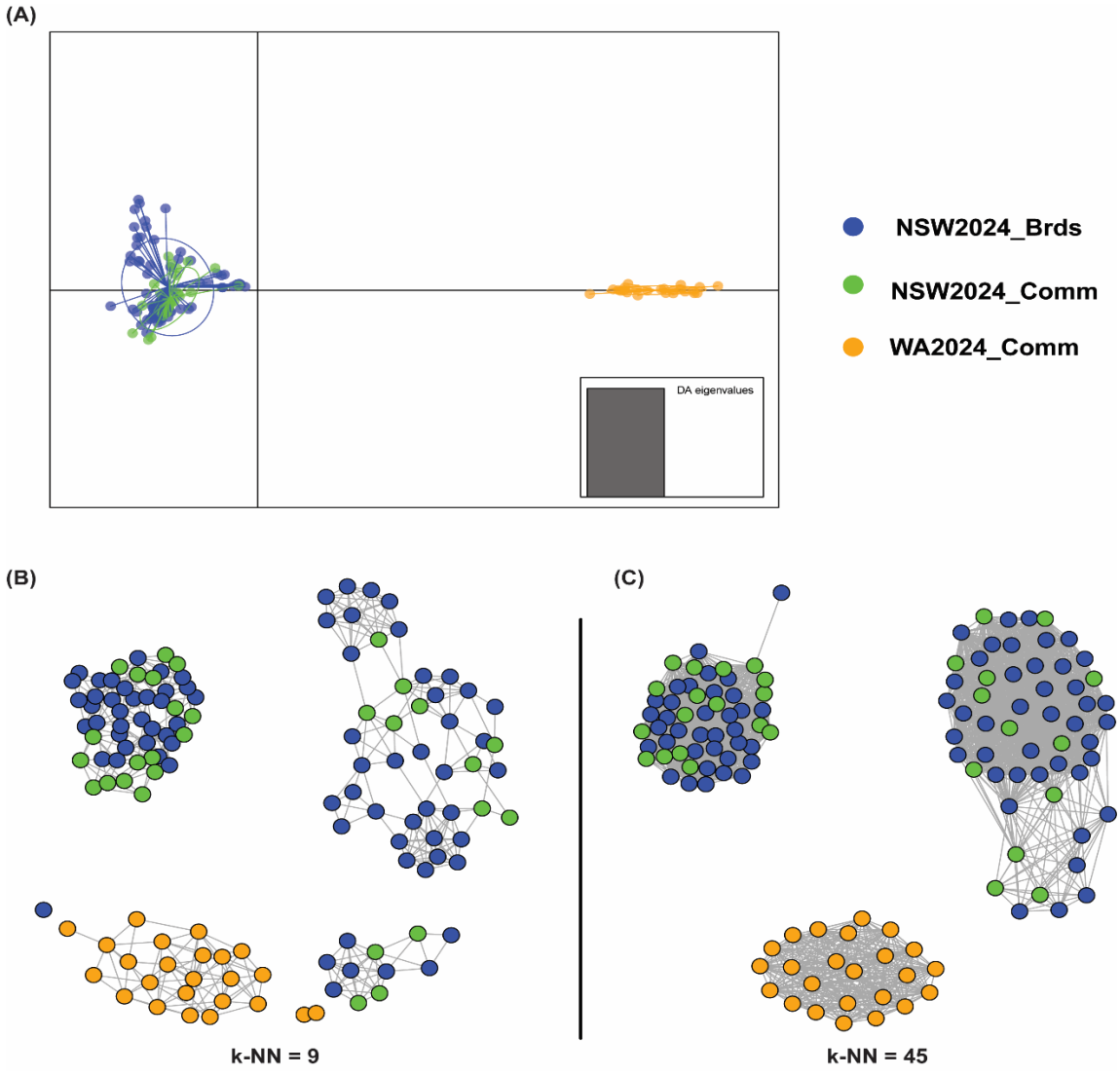

Figure 6 Population structure of Akoya oyster *P. fucata* sampled (n = 131 individuals) using 5,660 SNPs. Scatter plot of DAPC showing the first two discriminant functions **(A).** Population networks constructed using Netview at k-NN = 9 **(B)** and k-NN = 45 **(C)**. Dots represent individuals, whereas colored ellipses correspond to sampling origin. Population names are the same as defined in **Table 1**.

The NetView analysis supported these results (**Figure 6B** and **6C**). The optimal network connectivity determined by the Walktrap community detection algorithm was k-NN = 9, which separated the samples into nine groups; however, all NSW individuals clustered together, and WA individuals also clustered together. When the network connectivity increased (k-NN = 45), the NSW oysters were further divided into two subgroups, while the WA oysters remained as a single cluster, resulting in three groups overall.

4. Discussion

The Akoya oyster is a high-value aquaculture species, widely recognised for its importance in marine pearl production and increasingly valued as an emerging seafood product. However, the lack of efficient, reliable, and affordable genotyping tools has hindered the widespread application of genetic improvement programs in the Australian Akoya oyster. To address this limitation, this study developed and evaluated the first custom Allegro targeted SNP panel based on SPET technology, which offers an efficient and cost-effective genotyping strategy. This tool 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 SNP panel developed in this study integrated SNPs identified from two complementary sequencing platforms, WGS and ddRAD sequencing, allowing the panel to capture both reduced-representation and genome-wide variation. This approach enhances genomic coverage, reduces dependency on multiple sequencing datasets, and ensures more consistent marker representation for downstream analyses in aquaculture research and breeding programs. Following a rigorous multi-step filtering and probe design process (**Figure 1**), 9,448 SNPs were successfully incorporated, represented by a total of 17,182 probes, providing redundancy to maximise genotyping reliability. These SNPs are evenly distributed across the 14 chromosomes, with an average genome-wide spacing of approximately 109 kbp, minimising marker clustering and ensuring broad genome coverage.

The custom Allegro SNP panel showed high technical performance, with 9,041 of 9,448 SNPs (95.7%) successfully genotyped across 190 individuals, and 8,856 loci (97.9%) being polymorphic, indicating that the SNP selection and filtering strategy effectively enriched for informative markers. Average sample and SNP call rates were also high (90.4% and 90.7%, respectively; **Figure 3B and 3C**), demonstrating reliable performance across both invasive

(tissue) and non-invasive (swab) sample types. In addition, comparison with WGS data showed generally high concordance, averaging approximately 91% at both sample- and SNP-levels, though slightly lower than the internal repeatability and reproducibility observed within the panel itself (~97%; **Figure 4**). This difference likely reflects inherent technical variation between genotyping platforms, such as allele capture efficiency and platform-specific biases. Following quality control, a total of 5,660 high-quality SNPs were retained for downstream analyses (**Table S2**). Studies in other aquaculture species, including giant black tiger prawn *Penaeus monodon* (Guppy et al., 2020; Vu et al., 2023), barramundi *Lates calcarifer* (Massault et al., 2025), leopard coral grouper *Plectropomus leopardus* and Japanese flounder *Paralichthys olivaceus* (Qu et al., 2024), and silver-lipped pearl oyster *Pinctada maxima* (Vu et al., 2026), indicated that optimally designed panels of similar or smaller size can achieve high genomic prediction accuracy while also enabling robust analyses of population genetic diversity and structure.

Building on its robust genotyping performance, the Allegro SNP panel can be effectively applied to key areas of genetic research and breeding in Australian Akoya oyster, including parentage assignment and genomic-informed breeding. Using a high-quality SNP subset (n = 1,165 SNPs), parentage analyses successfully assigned both parents for 80% of progeny, while a small proportion of progeny could only be assigned to a single candidate parent. This reduced assignment rate likely reflects missing genotype data among the broodstock, as only 73 of 86 candidate parents were successfully genotyped due to post-spawning mortality, failure to open during the swabbing procedure, or DNA quality issues. Genomic relatedness among the 73 broodstock indicated that 28 individuals were involved in at least one pairwise comparison with elevated relatedness (GRM > 0.40), and a substantial number showed moderate relatedness (GRM 0.20-0.38), reflecting shared ancestry within the broodstock. These findings highlight the importance of using genomic information to guide mating designs and minimize inbreeding and demonstrate that the custom Allegro SNP array is well-suited for parentage assignment and genomic-informed breeding applications.

Population genetic analyses using high-quality SNPs of custom SNP panel (n = 5,660 SNPs) provided insights into the genetic diversity of Australian Akoya oyster aquaculture populations. Genetic diversity metrics were broadly similar among the NSW populations, whereas the WA population showed slightly lower polymorphism and allelic richness, although heterozygosity levels remained comparable, possibly reflecting the smaller sample size for WA. Estimates of N_e were relatively small for the NSW broodstock and commercial populations,

suggesting that only a limited number of individuals contributed disproportionately to the next generation. Such reduced Ne is commonly observed in mass-spawning oyster aquaculture systems and may result from unequal family contributions or sweepstakes reproductive success during spawning and larval rearing (Vu et al., 2026).

Population structure analyses indicated the presence of two distinct genetic stocks corresponding to the east coast (NSW) and west coast (WA) populations (**Figure 6**). Pairwise F_{ST} values between NSW and WA populations were relatively high (F_{ST} = 0.15-0.17), suggesting substantial genetic differentiation. Similar east–west population structure has been reported in several Australian marine species, including the giant black tiger prawn *P. monodon* (Vu et al., 2020), Australian Spanish mackerel *Scomberomorus commerson* (Buckworth et al., 1998), mud crab *Scylla serrata* (Gopurenko & Hughes, 2002), barramundi *L. calcarifer* (Loughnan et al., 2019), and blue swimmer crab *Portunus armatus* and *P. pelagicus* (Premachandra et al., 2023). Such patterns are often attributed to biogeographic and oceanographic barriers that limit gene flow between eastern and western Australian coasts.

Interestingly, both NetView and DAPC analyses suggested the presence of at least two genetic subgroups within the NSW aquaculture populations, with broodstock (wild-caught) and commercial farm individuals distributed across both clusters. This contrasts with previous studies of the Sydney rock oyster *Saccostrea glomerata*, which reported genetic homogeneity across eastern Australian populations using neutral SNPs (O'Hare et al., 2021). This difference likely reflects the longer aquaculture history of Sydney rock oysters along the eastern coast, where the exchange of stocks among farms may have promoted genetic homogenization, whereas Akoya oyster aquaculture in the region is more recent and limited. To fully characterise the population genetics of Australian Akoya oysters for fishery and aquaculture management, additional sampling of wild populations from multiple locations along both eastern and western coasts will be necessary. Overall, the custom Allegro SNP array developed in this study demonstrates strong utility for population genetic analyses in this species.

5. Limitations and future perspectives

Despite its robust performance, the custom Allegro SNP panel has some limitations. First, it was designed using the Chines Akoya oyster *P. fucata martensii* genome (Zheng, 2023), which may have led to ~10% of SNPs showing multiple alleles and ~15% exhibiting call rates below 90% due to sequence mismatches with the Australian genome. Second, the panel was

developed and tested primarily on aquaculture populations from a limited number of locations, so population-specific or rare variants may be underrepresented. Third, the sample sizes used to validate the SNP panel were relatively small, limiting analyses of genomic selection, including the estimation of genetic parameters, genomic predictions, and GWAS.

Future applications of the panel could include broader sampling of wild populations from multiple locations along both eastern and western coasts to fully characterise the population genetics of Australian Akoya oysters for fishery and aquaculture management. In addition, genomic selection studies should be conducted to evaluate the panel's utility for breeding programs. Finally, functional SNPs included in the array from previous studies have not yet been validated; future work should confirm their functional relevance and integrate them into selective breeding strategies

Data Availability

All data analysed in this report are available exclusively to the ITRH and the FA CRC members.

References

Adamkewicz, S. L., & Harasewych, M. (1996). Systematics and biogeography of the genus *Donax* (Bivalvia: Donacidae) in eastern North America. *American Malacological Bulletin*, *13*, 97-104.

Buckworth, R., Newman, S., Ovenden, J., Lester, R., & McPherson, G. (1998). The stock structure of northern and western Australian Spanish Mackerel. Final Report. *Fisheries Research and Development Corporation Project*, *159*, 225.

Chung, W. H., Howieson, J., & Chaklader, M. R. (2021). The ameliorative effects of low-temperature pasteurization on physicochemical and microbiological quality of raw Akoya pearl oyster (*Pinctada fucata*). *Food Control*, *129*, 108241.

Cropp, D., Koltasz, C., Boschetti, P., & Davidson, M. (2011). *Develop the non-maxima pearl industry at the Abrolhos Islands (Pinctada Imbricata)* (FRDC Project No: 2007/216). L. F. P. Ltd.

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.

Do, C., Waples, R. S., Peel, D., Macbeth, G., Tillett, B. J., & Ovenden, J. R. (2014). NeEstimator v2: re-implementation of software for the estimation of contemporary effective population size (Ne) from genetic data. *Molecular Ecology Resources*, *14*(1), 209-214. https://doi.org/https://doi.org/10.1111/1755-0998.12157

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.

Flanagan, S. P., & Jones, A. G. (2019). The future of parentage analysis: From microsatellites to SNPs and beyond. *Molecular Ecology*, *28*(3), 544-567.

Foll, M. (2012). Bayescan v2. 1 user manual. *Ecology*, *20*, 1450-1462. http://cmpg.unibe.ch/software/BayeScan/files/BayeScan2.1_manual.pdf

Foll, M., & Gaggiotti, O. (2008). A genome-scan method to identify selected loci appropriate for both dominant and codominant markers: a Bayesian perspective. *Genetics*, *180*(2), 977-993. https://doi.org/10.1534/genetics.108.092221

Garrison, E., & Marth, G. (2012). Haplotype-based variant detection from short-read sequencing. *arXiv preprint arXiv:1207.3907*.

Gopurenko, D., & Hughes, J. M. (2002). Regional patterns of genetic structure among Australian populations of the mud crab, *Scylla serrata* (Crustacea: Decapoda): evidence from mitochondrial DNA. *Marine and Freshwater Research*, *53*(5), 849-857.

Goudet, J. (2005). Hierfstat, a package for R to compute and test hierarchical F-statistics. *Molecular ecology notes*, *5*(1), 184-186.

Guppy, J. L., Jones, D. B., Kjeldsen, S. R., Le Port, A., Khatkar, M. S., Wade, N. M., Sellars, M. J., Steinig, E. J., Raadsma, H. W., & Jerry, D. R. (2020). Development and validation of a RAD-Seq target-capture based genotyping assay for routine application in advanced black tiger shrimp (*Penaeus monodon*) breeding programs. *BMC genomics*, *21*(1), 1-16.

Houston, R. D., Bean, T. P., Macqueen, D. J., Gundappa, M. K., Jin, Y. H., Jenkins, T. L., Selly, S. L. C., Martin, S. A., Stevens, J. R., & Santos, E. M. (2020). Harnessing genomics to fast-track genetic improvement in aquaculture. *Nature Reviews Genetics*, *21*(7), 389-409.

Jerry, D. R., Jones, D. B., Lillehammer, M., Massault, C., Loughnan, S., Cate, H. S., Harrison, P. J., Strugnell, J. M., Zenger, K. R., & Robinson, N. A. (2022). Predicted strong genetic gains from the

application of genomic selection to improve growth related traits in barramundi (*Lates calcarifer*). *Aquaculture*, *549*, 737761.

Jombart, T. (2008). adegenet: a R package for the multivariate analysis of genetic markers. *Bioinformatics*, *24*(11), 1403-1405. https://doi.org/10.1093/bioinformatics/btn129

Jombart, T., & Ahmed, I. (2012). Adegenet: an R package for the exploratory analysis of genetic and genomic data. *R package 'Adegenet' version*, 1.3-7.

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

Keenan, K., McGinnity, P., Cross, T. F., Crozier, W. W., & Prodöhl, P. A. (2013). diveRsity: An R package for the estimation and exploration of population genetics parameters and their associated errors. *Methods in Ecology and Evolution*, *4*(8), 782-788. https://doi.org/https://doi.org/10.1111/2041-210X.12067

Knaus, B. J., & Grünwald, N. J. (2017). vcfr: a package to manipulate and visualize variant call format data in R. *Molecular Ecology Resources*, *17*(1), 44-53.

Kriaridou, C., Tsairidou, S., Fraslin, C., Gorjanc, G., Looseley, M. E., Johnston, I. A., Houston, R. D., & Robledo, D. (2023). Evaluation of low-density SNP panels and imputation for cost-effective genomic selection in four aquaculture species. *Frontiers in genetics*, *14*, 1194266.

Kuchel, R. P., O'Connor, W. A., & Raftos, D. A. (2011). Environmental stress and disease in pearl oysters, focusing on the Akoya pearl oyster (*Pinctada fucata* Gould 1850). *Reviews in Aquaculture*, *3*(3), 138-154.

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, H. (2013). Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv preprint arXiv:1303.3997*.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., & Durbin, R. (2009). The sequence alignment/map format and SAMtools. *Bioinformatics*, *25*(16), 2078-2079.

Lin, W., Xiao, Q., Yu, F., Han, Z., Liu, J., Peng, W., Huang, Z., Lei, Y., Li, W., & You, W. (2023). Development of a low-density SNP genotyping panel by a novel technology mGPS and its application in germplasm identification of abalone. *Aquaculture*, *565*, 739089.

Liu, H., Zhang, H., Pan, X., Xu, M., Huang, J., & He, M. (2020). A high density genetic map by whole-genome resequencing for QTL fine-mapping and dissecting candidate genes for growth or sex traits in the pearl oyster (*Pinctada fucata martensii*). *Aquaculture*, *519*, 734839.

Loughnan, S. R., Smith-Keune, C., Beheregaray, L. B., Robinson, N. A., & Jerry, D. R. (2019). Population genetic structure of barramundi (*Lates calcarifer*) across the natural distribution range in Australia informs fishery management and aquaculture practices. *Marine and Freshwater Research*, -. https://doi.org/https://doi.org/10.1071/MF18330

Luu, K., Bazin, E., & Blum, M. G. (2017). pcadapt: an R package to perform genome scans for selection based on principal component analysis. *Molecular Ecology Resources*, *17*(1), 67-77. https://doi.org/10.1111/1755-0998.12592

Massault, C., Jeffrey, C., Jones, D. B., Barnard, R., Strugnell, J. M., Zenger, K. R., & Jerry, D. R. (2022). Non-invasive DNA collection for parentage analysis for bivalves: A case study from the silver-

lipped pearl oyster (*Pinctada maxima*). *Aquaculture*, *552*, 738036. https://doi.org/https://doi.org/10.1016/j.aquaculture.2022.738036

Massault, C., Jones, D., Zenger, K., Strugnell, J., Barnard, R., & Jerry, D. (2021). A SNP parentage assignment panel for the silver lipped pearl oyster (*Pinctada maxima*). *Aquaculture Reports*, *20*, 100687. https://doi.org/https://doi.org/10.1016/j.aqrep.2021.100687

Massault, C., Jones, D. B., Lillehammer, M., Cate, H., Harrison, P., Strugnell, J. M., Zenger, K. R., Robinson, N. A., & Jerry, D. R. (2025). Accuracies of genomic prediction accounting for genotype by environment remain high when using small sets of selected SNPs in barramundi *Lates calcarifer*. *Aquaculture*, *599*, 742138.

Matsuyama, T., Miwa, S., Mekata, T., Matsuura, Y., Takano, T., & Nakayasu, C. (2021). Mass mortality of pearl oyster (*Pinctada fucata* (Gould)) in Japan in 2019 and 2020 is caused by an unidentified infectious agent. *PeerJ*, *9*, e12180.

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., & Daly, M. (2010). The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome research*, *20*(9), 1297-1303.

Mijangos, J. L., Gruber, B., Berry, O., Pacioni, C., & Georges, A. (2022). dartR v2: An accessible genetic analysis platform for conservation, ecology and agriculture. *Methods in Ecology and Evolution*, *13*(10), 2150-2158. https://doi.org/https://doi.org/10.1111/2041-210X.13918

Miyazaki, T., Goto, K., Kobayashi, T., Kageyama, T., & Miyata, M. (1999). Mass mortalities associated with a virus disease in Japanese pearl oysters *Pinctada fucata martensii*. *Diseases of aquatic organisms*, *37*, 1-12.

Nagai, K. (2013). A history of the cultured pearl industry. *Zoological science*, *30*(10), 783-793.

Neuditschko, M., Khatkar, M. S., & Raadsma, H. W. (2012). NetView: a high-definition network-visualization approach to detect fine-scale population structures from genome-wide patterns of variation. *PloS one*, *7*(10), e48375. https://doi.org/10.1371/journal.pone.0048375

O'Connor, W., Lawler, N., & Heasman, M. (2003). Trial farming the akoya pearl oyster, Pinctada imbricata. *Port Stephens, NSW. NSW Fisheries Final Report Series*, *42*, 174.

O'Hare, J. A., Momigliano, P., Raftos, D. A., & Stow, A. J. (2021). Genetic structure and effective population size of Sydney rock oysters in eastern Australia. *Conservation Genetics*, *22*(3), 427-442.

Peterson, B. K., Weber, J. N., Kay, E. H., Fisher, H. S., & Hoekstra, H. E. (2012). Double digest RADseq: an inexpensive method for de novo SNP discovery and genotyping in model and non-model species. *PloS one*, *7*(5), e37135.

Pit, J. H. (2004). *Feasibility of Akoya oyster culture in Queensland* James Cook University]. Townsville, Australia.

Premachandra, H., Becker, A., Millard, K., Johnston, D., Subramanian, S., Kumar, M., Taylor, M., & Knibb, W. (2023). Genomic analyses indicate two blue swimmer crab species in Australia, evidence for natural interspecific hybridization and genetic structure within species with implications for fisheries management and stock enhancement. *Fisheries Research*, *265*, 106757.

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A., Bender, D., Maller, J., Sklar, P., De Bakker, P. I., & Daly, M. J. (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. *The American Journal of Human Genetics*, *81*(3), 559-575. https://doi.org/https://doi.org/10.1086/519795

Qu, S., Lu, S., Liu, Y., Li, M., & Chen, S. (2024). Accurate genomic selection using low-density SNP panels preselected by maximum likelihood estimation. *Aquaculture*, *579*, 740154.

Robledo, D., Palaiokostas, C., Bargelloni, L., Martínez, P., & Houston, R. (2018). Applications of genotyping by sequencing in aquaculture breeding and genetics. *Reviews in Aquaculture*, *10*(3), 670-682.

Scaglione, D., Pinosio, S., Marroni, F., Di Centa, E., Fornasiero, A., Magris, G., Scalabrin, S., Cattonaro, F., Taylor, G., & Morgante, M. (2019). Single primer enrichment technology as a tool for massive genotyping: a benchmark on black poplar and maize. *Annals of botany*, *124*(4), 543-551.

Steinig, E. J., Neuditschko, M., Khatkar, M. S., Raadsma, H. W., & Zenger, K. R. (2016). netview p: a network visualization tool to unravel complex population structure using genome-wide SNPs. *Molecular Ecology Resources*, *16*(1), 216-227. https://doi.org/https://doi.org/10.1111/1755-0998.12442

Takeuchi, T., Suzuki, Y., Watabe, S., Nagai, K., Masaoka, T., Fujie, M., Kawamitsu, M., Satoh, N., & Myers, E. W. (2022). A high-quality, haplotype-phased genome reconstruction reveals unexpected haplotype diversity in a pearl oyster. *DNA research*, *29*(6), dsac035.

Theodorou, J. A., Bekiari, V., Douvi, X., Ramfos, A., & Tzovenis, I. (2024). Analysis of nutritional composition of the pearl oyster Pinctada radiata as a new Mediterranean "bioresource" for human consumption. *Applied Sciences*, *14*(21), 9757.

Vu, N. T., Jerry, D. R., Edmunds, R. C., Jones, D. B., & Zenger, K. R. (2023). Development of a global SNP resource for diversity, provenance, and parentage analyses on the Indo-Pacific giant black tiger shrimp (*Penaeus monodon*). *Aquaculture*, *563*, 738890.

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

Vu, N. T. T., Zenger, K. R., Guppy, J. L., Sellars, M. J., Silva, C. N. S., Kjeldsen, S. R., & Jerry, D. R. (2020). Fine-scale population structure and evidence for local adaptation in Australian giant black tiger shrimp (*Penaeus monodon*) using SNP analysis. *BMC genomics*, *21*(1), 669. https://doi.org/10.1186/s12864-020-07084-x

Wada, K. T., & Tëmkin, I. (2008). Taxonomy and phylogeny. *The pearl oyster*, *2008*, 37-75.

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.

Weir, B. S., & Cockerham, C. C. (1984). Estimating F-statistics for the analysis of population structure. *Evolution*, *38*(6), 1358-1370. https://doi.org/10.2307/2408641

Wenne, R. (2018). Single nucleotide polymorphism markers with applications in aquaculture and assessment of its impact on natural populations. *Aquatic Living Resources*, *31*, 2.

Wickham, H., Chang, W., Henry, L., Pedersen, T., Takahsahi, K., Wilke, C., Woo, K., & Yutani, H. (2016). ggplot2: Elegant Graphics for Data Analysis (version 3.2. 1). In: Springer-Verlag New York.

Wickham, H., Pedersen, T., & Seidel, D. (2025). *scales: Scale Functions for Visualization*. In (Version 1.4.0) R package. https://scales.r-lib.org, https://github.com/r-lib/scales

Wright, S. (1949). The genetical structure of populations. *Annals of eugenics*, *15*(1), 323-354. https://doi.org/10.1136/bmj.2.4669.36

Wu, Y., Wang, J., Li, L., Yang, X., Wang, J., & Hu, X. (2018). Purification and identification of an antioxidant peptide from *Pinctada fucata* muscle. *CyTA-Journal of Food*, *16*(1), 11-19.

Yang, B., Li, Y., Xu, M., Li, Q., & Liu, S. (2025). Developing efficient and cost-effective NGS sequencing libraries for high throughput genotyping by whole genome resequencing in aquaculture. *Aquaculture*, 743431.

Zenger, K. R., Khatkar, M. S., Jones, D. B., Khalilisamani, N., Jerry, D. R., & Raadsma, H. W. (2019). Genomic selection in aquaculture: application, limitations and opportunities with special reference to marine shrimp and pearl oysters. *Frontiers in genetics*, *9*, 693. https://doi.org/https://doi.org/10.3389/fgene.2018.00693

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.

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.

Supplementary Information

Table S1 SNP calling and filtering steps to identify high-quality SNPs for the custom SNP panel.

SNP calling - filtering	Whole genome sequencing (WGS)			ddRAD-NSW		ddRAD-WA	
	Freebayes	**Samtools**	**gatk**	**Freebayes**	**Samtools**	**Freebayes**	**Samtools**
Initial SNPs	**73,918,199**	**88,932,218**	**90,146,479**	**1,623,795**	**2,537,449**	**2,618,021**	**3,396,402**
Number of Akoya samples (individuals)	**20**	**20**	**20**	**99**	**99**	**115**	**115**
Hard quality control							
1. Sequence quality (QUAL score < 20)	**53,181,942** (Removed 20,736,257)	**78,511,175** (Removed 10,421,043)	**90,146,479** (Removed 0)	**1,393,870** (Removed 229,925)	**2,026,507** (Removed 510,924)	**2,307,386** (Removed 310,635)	**2,932,455** (Removed 463,947)
2. Read depth and missing data per SNP (Mean read depth per sample < 5 or > 25 for WGS, and < 3 or > 25 for ddRAD; missing data per SNP ≥ 10%)	**36,428,993** (Removed 16,752,949)	**50,177,865** (Removed 28,333,310)	**70,058,772** (Removed 20,087,707)	**99,480** (Removed 1,194,910)	**438,626** (Removed 1,587,881)	**58,638** (Removed 2,248,748)	**700,313** (Removed 2,232,142)
3. Chromosome filtering SNPs located outside the 14 assembled chromosomes were removed	**36,372,868** (Removed 56,125)	**48,424,089** (Removed 78,402)	**69,914,964** (Removed 143,808)	**99,415** (Removed 65)	**437,866** (Removed 760)	**58,600** (Removed 38)	**699,163** (Removed 1,150)
4. Variant type Indels removed (SNPs retained only)	**30,783,368** (Removed 5,589,500)	**48,424,089** (Removed 0)	**57,731,939** (Removed 12,183,025)	**87,700** (Removed 11,715)	**410,331** (Removed 27,535)	**50,493** (Removed 8,170)	**646,202** (Removed 52,961)
5. Merging for common	**23,591,082**			**78,179**		**46,761**	

Number of Akoya samples (individuals)	**19**	**99**	**109**
SNP quality control			
6. Removed multiallelic	**19,658,614** (Removed 3,932,468)	**68,132** (Removed 10,047)	**41,657** (Removed 5,104)
7. Removed in low-complexity and soft masked regions	**5,626,477** (Removed 14,032,137)	**24,511** (Removed 43,621)	**15,928** (Removed 25,729)
8. Removed monomorphic and near indels	**4,910,198** (Removed 716,279)	**23,526** (Removed 985)	**15,127** (Removed 801)
9. Removed with MAF ≤ 0.05	**3,957,137** (Removed 953,061)	**12,104** (Removed 11,422)	**10,444** (Removed 4,683)
10. Remove deviating from HWE ($p < 0.01$)	**3,937,263** (Removed 19,874)	**9,369** (Removed 2,735)	**8,586** (Removed 1,858)
11. LD pruning (retain one SNP per pair, $r^2 > 0.2$)	**341,239** (Removed 3,596,024)	**4,017** (Removed 5,352)	**2,191** (Removed 6,395)

* NSW: New South Wales; WA: Western Australia; ddRAD: double-digest RAD sequencing; MAF: minor allele frequencies; HWE: Hardy-Weinberg Equilibrium; LD: linkage disequilibrium.

Table S2: Summary of SNP filtering steps for validation of the custom Allegro SNP panel.

SNP quality control step	Number of SNPs removed	Remaining SNPs	Notes
Raw SNPs from the custom Allegro panel	-	9,041	SNPs successfully genotyped
Sequence quality (QUAL score ≥ 20)	0	9,041	High confidence variants (≥ 99% probability)
Indels	296	8,745	Removed indels, retained SNPs only
Multiallelic SNPs	926	7,819	Retained only biallelic SNPs
SNPs with call rate ≥ 90%	1,379	6,422	Removed SNPs with low genotyping success
SNPs failing MAF ≥ 0.02	284	6,138	Retained SNPs with MAF ≥ 0.02
LD pruning ($r^2 > 0.2$)	478	5,660	Removed one SNP from each pair with $r^2 > 0.2$
Final SNPs retained	-	**5,660**	High-quality SNPs for downstream analyses

* QUAL: Phred-scaled probability score; MAF: Minor Allele Frequency; LD: linkage disequilibrium

Table S3 Sample and SNP-level genotype concordance between duplicated sample pairs.

A. Sample-level genotype concordance

Comparison	No of sample pairs	Min (%)	Max (%)	Mean ± SD (%)
Between duplicated samples in WGS vs Allegro panel	19	86.65	92.81	91.29 ± 1.38
Between duplicated samples within a same genotype batch	6	96.86	97.91	97.37 ± 0.32
Between duplicated samples across two genotype batches	6	96.20	97.03	96.6 ± 0.30

B. SNP-level genotype concordance

Comparison	No of SNPs	Min (%)	Max (%)	Mean ± SD (%)
Between duplicated samples in WGS vs Allegro panel	5,519	10.53	100	90.85 ± 14.34
Between duplicated samples within a same genotype batch	6,208	16.67	100	97.28 ± 8.30
Between duplicated samples across two genotype batches	6,085	16.67	100	96.52 ± 9.17

Table S4 Population genetic differentiation estimated for aquaculture Akoya oyster *Pinctada fucata* sampled.

	NSW2024_Brds	NSW2024_Comm	WA2024_Comm
NSW2024_Brds	0		
NSW2024_Comm	0.001	0	
NSW2024_Comm	0.15	0.17	0

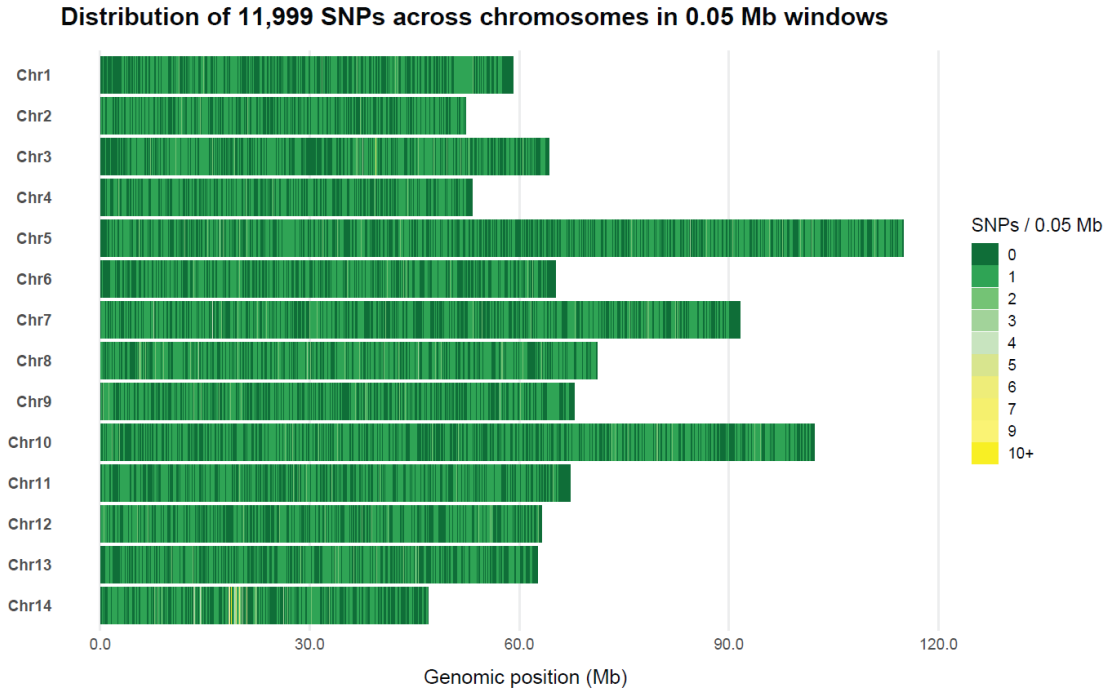

Figure S1 Genome-wide distribution of the 11,999 SNPs probe design for the custom Allegro SNP panel across 14 chromosomes. SNP positions were divided into non-overlapping 50 kbp (0.05 Mb) windows. Tile colour indicates the number of SNPs per window, with counts capped at 10.

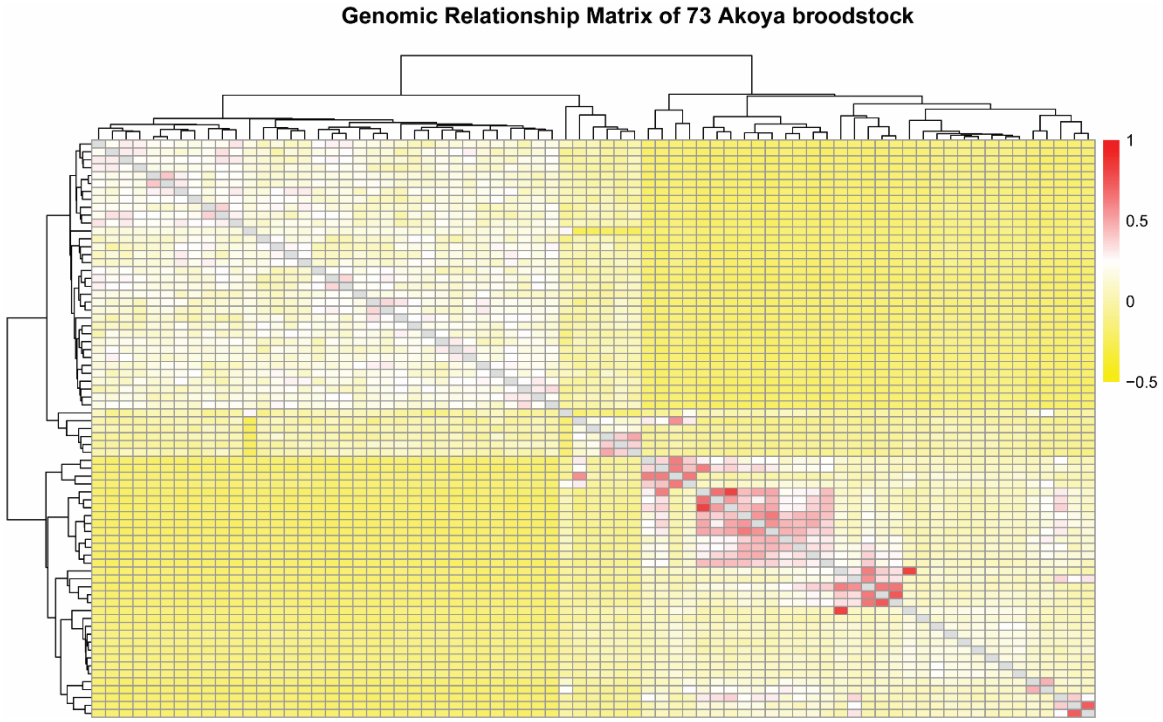

Figure S2 Genomic relationship matrix (GRM) showing pairwise genomic relatedness among 73 Akoya oyster broodstock *Pinctada fucata* based on 5,660 SNPs. The heatmap was generated using hierarchical clustering of individuals. Colours represent genomic relationship coefficients ranging from −0.5 (yellow) to 1 (red), with white indicating intermediate relatedness.