An amplicon panel for high-throughput and low-cost genotyping of Pacific oyster

Ben J. G. Sutherland[1,2], Neil F. Thompson[3], Krishna Reddy Gujjula[4], Claudio D. Carrasco[4], Srinivas Chadaram[4], Amy M. Chan[5], Curtis A. Suttle[5,6,7,8], Timothy J. Green[2]

[1] Sutherland Bioinformatics, Lantzville, BC, Canada V0R 2H0

[2] Faculty of Science and Technology, Vancouver Island University, Nanaimo, BC, Canada V9R 5S5

[3] Pacific Shellfish Research Unit, Agricultural Research Service, United States Department of Agriculture, Hatfield Marine Science Center, Newport, OR 97365 USA

[4] ThermoFisher Scientific, 2130 Woodward Street, Austin, TX 78744 USA

[5] Department of Earth, Ocean and Atmospheric Sciences, The University of British Columbia, Vancouver, BC, Canada V6T 1Z4

[6] Department of Microbiology and Immunology, The University of British Columbia, Vancouver, BC, Canada V6T 1Z3

[7] Department of Botany, The University of British Columbia, Vancouver, BC, Canada V6T 1Z4

[8] Institute for the Oceans and Fisheries, The University of British Columbia, Vancouver, BC, Canada, V6T 1Z4

*Author for correspondence: BJGS

Current Address:

Sutherland Bioinformatics, Lantzville BC, Canada, V0R 2H0

Sutherland.Bioinformatics@protonmail.com

Running Title: *Pacific oyster amplicon panel*

Abstract

Maintaining genetic diversity in cultured shellfish can be challenging due to high variance in individual reproductive output, founder effects, and rapid genetic drift. However, it is important to consider for retaining adaptive potential and avoiding inbreeding depression, particularly in species with high mutational load. To support management of genetic diversity and selective breeding efforts in cultured Pacific oyster (*Crassostrea gigas*; a.k.a. *Magallana gigas*), we developed an amplicon panel comprising 592 amplicons positioned throughout the genome with an average of 50 markers per chromosome. Markers were selected based on high heterozygosity or high differentiation in Pacific oyster populations from the eastern Pacific Ocean. The developed panel was tested in a pilot study on 371 unique individuals, including those from hatcheries and farms (n = 72) and from three generations of the Vancouver Island University (VIU) breeding program (n = 181), as well as from naturalized or wild populations in the Northern Hemisphere (n = 118). High repeatability was observed in technical replicate samples; on average, 97.5% of typed markers were concordant between samples (n = 68 replicate sample pairs). In the pilot dataset, following quality filters, 463 markers were retained for analysis (average observed heterozygosity = 0.29 ± 0.19; and F_{ST} = 0.040 ± 0.046). Principal components analysis showed separate clustering of the Japan translocation lineage, China, and a commercial aquaculture population. Parentage analysis of the VIU breeding program identified 84% and 66% of the possible parent-offspring assignments in two spawning events involving 98 and 40 offspring each, respectively. Results were strongest in individuals with the highest genotyping rates, highlighting the value of sample and target SNP quality. The analysis identified pedigree errors that were then corrected. All code required to analyze the data is provided, including *amplitools*, a repository to move the panel target SNP output through parentage analysis. Collectively, these advances will enable rapid, consistent, and cost-effective genotyping of Pacific oyster.

Keywords: amplicon panel, aquaculture, *Crassostrea gigas*, genetic diversity, *Magallana gigas*, Pacific oyster, parentage

Introduction

Sustainable aquaculture depends on the effective characterization, preservation, exchange and use of genetic resources (Guo 2009). High resolution genomic tools, such as high-density genetic maps, SNP chips, and chromosome-level reference genomes enable genome-wide association studies (GWAS), the identification of quantitative trait loci (QTL), and marker-assisted or genomic selection (Boudry *et al.* 2021). Additionally, lower resolution tools such as low or medium-density marker panels also have significant value, enabling high-throughput genotyping at low cost. Low-density panels can provide applied benefits to fisheries management and aquaculture production; genotyping all broodstock of salmon hatcheries inherently tags offspring for future detection in fisheries (e.g., Beacham *et al.* 2018), and enables common-garden evaluation of breeding program cohorts for selection traits by identifying family-of-origin without the need for physical tags (Allen *et al.* 2020). Amplicon panels are useful for genotyping with rapid turnaround time (Arbelaez *et al.* 2019), are reliable, scalable, fairly easy to analyze, and cost-effective (Meek & Larson 2019), and can be more robust to lower quality DNA than high-density methods (Csernák *et al.* 2017). The reduced cost of low-density methods can bring genetic advances to small or medium sized operations that are unable to afford high-density genotyping tools (Boudry *et al.* 2021; Delomas *et al.* 2023).

Pacific oyster *Crassostrea gigas* (a.k.a. *Magallana gigas*; Salvi & Mariottini 2017) is one of the most valuable aquaculture species globally (Botta *et al.* 2020). Oyster aquaculture is a significant economic driver and food source for many developed and developing nations (Martínez-García *et al.* 2022). However, production and growth of the industry outside of China has stagnated due to disease, regulatory issues, and other factors (Botta *et al.* 2020). Large-scale mortality events have been a significant burden on the industry, where disease may be caused by specific or combined protozoan, bacterial, or viral pathogens, as well as by multifactorial or unexplained causes (King *et al.* 2019). Abiotic stressors such as salinity shifts, elevated temperatures, or heat waves can interact with these biotic stressors, increasing negative impacts (Green *et al.* 2019; King *et al.* 2019). Environmental changes more generally are also challenging cultured and wild oysters. Ocean acidification negatively impacts shell formation and maintenance (Watson *et al.* 2009; Barton *et al.* 2012; Barros *et al.* 2013), and extreme weather has led to massive mortality events in recent years. For example, an atmospheric river and freshwater influx into the San Francisco Bay (California, USA) coincided with a near 100% mortality event of the highly abundant population of wild Olympia oyster *Ostrea lurida* (Cheng *et al.* 2016). An unprecedented extreme heatwave in the Pacific Northwest during summer 2021 resulted in large-scale negative impacts to wild and naturalized shellfish in the area (Raymond *et al.* 2022). With a multitude of stressors and impediments to industry stability and growth, shellfish farmers in the Pacific Northwest consider oyster health and selective breeding to be important scientific adaptation strategies to increase crop resilience (Green *et al.* 2023). As such, improvements to selective

breeding infrastructure and genomic resources are valuable to Pacific oyster, as recently advocated for eastern oyster *C. virginica* (Allen *et al.* 2020).

Maintenance of genetic diversity is critical in shellfish breeding. Although polymorphism rates are very high in shellfish (Plough 2016), where a SNP is expected every 40 bp in *C. gigas* (Hedgecock *et al.* 2005; Sauvage *et al.* 2007), rapid reductions in genetic diversity can occur in cultured lineages (Hedgecock & Sly 1990; Evans *et al.* 2004; Xiao *et al.* 2011; Gurney-Smith *et al.* 2017), with losses compounding over time, leading to significant concerns for breeders (e.g., Li *et al.* 2007). The high polymorphism rate also results in high genetic load, which can produce severe inbreeding depression affecting growth and survival (Evans *et al.* 2004). Although inbreeding depression may be counteracted by crossing divergent lines to achieve heterosis (Hedgecock *et al.* 1995; Hedgecock & Davis 2007), long-term retention of genetic diversity, and maintenance of hatchery lineages is important for broodstock programs (Carlsson *et al.* 2006; Gurney-Smith *et al.* 2017). Furthermore, retaining standing genetic diversity is vital for rapid adaptation (Barrett & Schluter 2008) and selective breeding (Guo 2009).

A rapid, inexpensive, automatable, and easily analyzed genotyping tool is needed for the Pacific oyster. Existing genomic resources for Pacific oyster include high quality reference genome assemblies (Peñaloza *et al.* 2021; Qi *et al.* 2021) and genetic maps (Gutierrez *et al.* 2018; Li *et al.* 2018), as well as SNP arrays of high (134K targets; Qi *et al.* 2017), medium (27K; Gutierrez *et al.* 2017), and low density (384; Lapègue *et al.* 2014). However, due to significant costs and required expertise, smaller or medium-scale breeding operations may not be able to benefit from these resources (Boudry *et al.* 2021). Until present, there has been no low-density amplicon panel for Pacific oyster. In addition to lower cost per individual, amplicon panels can also genotype non-target variants within an amplicon window; this may expand the utility across populations and time, as long as primer sites remain functional. This longevity and malleability may be particularly advantageous to Pacific oyster given its high genetic diversity (Hedgecock *et al.* 2005; Sauvage *et al.* 2007), the global scale of their culture (Boudry *et al.* 2021; Martínez-García *et al.* 2022) including genetically diverse large-scale commercial hatcheries (Sutherland *et al.* 2020), and the temporal fluctuation in allele frequencies due to sweepstakes reproductive success (Hedgecock 1994; Hedgecock & Pudovkin 2011).

Here we describe the development of a 592 marker panel based on high heterozygosity and high differentiation markers from Pacific oyster populations in in British Columbia, Canada, that is documented to be similar to other populations derived from the Japan translocation lineage (Sutherland *et al.* 2020) and tested the panel on Pacific oysters from Canada, France, Japan, and China, as well as cultured populations from the United Kingdom and Canada. We also evaluated and demonstrated the panel's effectiveness for genetic parentage assignment using single SNP targets through analyzing two sequential spawning events (i.e., F0, F1, and F2 generations) from the Vancouver Island University (VIU) breeding program. All design

amplicons and target variants are provided here, and the panel can be obtained through a commercial provider. All analytic code is available in *amplitools*, a version-controlled repository designed to move from sequencer output to parentage analysis. Details on the latest version of the panel's target file is available in the accompanying repository *amplitargets*, expected to be updated over time. Collectively, with these outputs we aim to facilitate uptake and therefore bring rapid gains to Pacific oyster breeding and aquaculture.

Methods

Data sources and marker selection

Plink and Variant Call Format (VCF) outputs from Sutherland et al. (2020), a restriction-site associated DNA sequencing (RAD-seq) study of Pacific oysters from the Northern Hemisphere, were obtained from the authors and uploaded to FigShare (see *Data Availability*). Genotypes were read into R (R Core Team 2023) using the read.PLINK function of adegenet (v.2.1.5; Jombart & Ahmed 2011) and data were converted from genlight format to genind, genepop, and hierfstat formats for analysis using adegenet and hierfstat (v.0.5-10; Goudet & Jombart 2022).

The pre-filtered dataset was limited to only include naturalized collections from British Columbia (BC), the focal region of the study, which includes Hisnit (HIS), Pendrell (PEN), Pipestem (PIP), and Serpentine; SER; for details on collection sites see Sutherland et al. (2020). Within the BC dataset, per locus global minor allele frequency (MAF) was recalculated, and markers with MAF < 0.01 were removed. Per locus observed heterozygosity (H_{OBS}) and average F_{ST} (Weir & Cockerham 1984) were calculated using adegenet and pegas (v.0.12; Paradis 2010), respectively. Markers were excluded when $H_{OBS} > 0.5$. Markers with the highest H_{OBS} (n = 300), or the highest F_{ST} (n = 300) were selected for panel design. Several additional markers of interest were included based on previous detection as private alleles in Deep Bay, BC (n = 15) or Guernsey, UK (n = 5) farmed populations (Sutherland *et al.* 2020). Nine markers were redundant between the F_{ST} and H_{OBS} selected lists; 611 unique markers were identified for panel design (Table 1).

Panel design and chromosomal locations

Flanking 200 bp sequences on either side of selected variants were obtained from the contig-level reference genome (GCA_000297895.1; Zhang *et al.* 2012) used in marker discovery (Sutherland *et al.* 2020) using a bed file derived from the marker discovery VCF and the bedtools function getfasta (Quinlan & Hall 2010). All required code and instructions for this process are provided in the repository *ms_amplicon_panel* (see *Data Availability*). Marker sequences (401 bp), variant positions, and reference and alternate alleles (Additional File S1) were submitted to the AgriSeq primer design team for the Ion Torrent system (ThermoFisher Scientific). The markers were passed though the AgriSeq quality control process

156 (ThermoFisher Scientific), using the GCA_000297895.1 genome. Markers passing QC were then submitted
157 to the primer design phase. In the design phase, oligo candidates were generated, scored, filtered, and the
158 most optimal oligo pairs for each marker were selected to be included in the final panel. The selected oligos
159 were checked *in silico* for specificity and sensitivity of the intended target/marker regions using the *C. gigas*
160 reference genome GCA_000297895.1. The resultant marker panel is available commercially through
161 ThermoFisher Scientific (SKU A58237 AGRISEQ PACIFIC OYSTER PANEL).
162 The RAD-seq marker discovery used the contig-level genome assembly (Zhang *et al.* 2012), and
163 so the chromosomal positions of the markers were determined by mapping the 401 bp target sequences
164 against a chromosome assembly (GCF_902806645.1; Peñaloza *et al.* 2021) using bowtie2 (Langmead &
165 Salzberg 2012) in end-to-end mode allowing multiple alignments (-k 6). Results were categorized as single
166 or multiple mapping markers. Positions of markers along chromosomes were plotted using ggplot2
167 (Wickham 2016) (see *Data Availability*).
168
169 *Panel testing*
170 The panel was tested by genotyping populations collected in 2017 and 2018, as described by Sutherland et
171 al. (2020), including Deep Bay (DPB; BC farm), Pendrell Sound (PEN; naturalized BC), Guernsey (GUR;
172 commercial hatchery, United Kingdom), France (FRA; naturalized), Japan (JPN; wild), and China (CHN;
173 wild; and QDC; commercial hatchery; Table 2). These same populations and collection years were used for
174 marker discovery (Sutherland *et al.* 2020), but the panel testing was done on alternate individuals not used
175 for discovery, except for JPN and CHN, as additional samples were not available. Notably, JPN and CHN
176 were not used for marker selection (*see above)*. Secondly, the panel was tested for utility in parentage
177 analysis by genotyping samples from three generations of the Vancouver Island University (VIU) breeding
178 program. The breeding program samples were provided with information about their respective generation,
179 but pedigree information was not used until after assignments were created and initial inspections conducted.
180 DNA from VIU breeding program samples was extracted using the Monarch Genomic DNA
181 Purification kit (NEB) using the enzymatic cleanup tissue extraction protocol with both proteinase K and
182 RNase A, and quantified using a BioSpectrometer (Eppendorf). All other samples were extracted using the
183 BioSprint (QIAGEN) method and diluted to 25 ng/µl and stored at -20ºC. Undiluted samples were then
184 normalized to 10 ng/µl in 25 µl volumes and submitted in microtitre plates to ThermoFisher Scientific
185 laboratories in Austin, TX for AmpliSeq library preparation using the AgriSeq™ HTS Library kit
186 (ThermoFisher Scientific) with 16 cycles of amplification and IonCode barcode labelling. Libraries were
187 sequenced using Ion 540 chips on an Ion Torrent S5 (ThermoFisher Scientific). Some samples had low
188 sequence yields, and so a portion of the poorly amplifying libraries were resynthesized and sequenced at
189 both diluted and supplied sample concentrations. Replicate samples with sufficient genotyping rates were

used to evaluate repeatability of library preparation, sequencing, and genotyping (*see below for details*). Markers were scored using the Torrent Suite software VariantCaller plugin (ThermoFisher Scientific) using a hotspot file (Cgig v.1.0 hotspots; see *Data Availability*). The output tab-delimited multi-locus genotypes files (see *Data Availability*) were filtered for hotspot variants only (i.e., the target variant identified in marker discovery), then were converted to genepop format using the custom pipeline *amplitools* (see *Data Availability)*.

Repeatability, filtering, and population genetic analyses

The pilot study genepop file, including all samples, was imported into R using adegenet as implemented in the custom pipeline *simple_pop_stats* using instructions provided in *ms_oyster_panel* (see *Data Availability*). Technical replicates were inspected using the function *comp_tech_reps* within *simple_pop_stats*. When at least two technical replicates were genotyped at more than 50% of the loci they were compared at the level of the genotype. A genind file was generated including only unique individuals, keeping the best replicate (i.e., the one with the highest genotyping rate) when more than one replicate of a single individual was present. Samples or markers with more than 50% of genotypes missing were removed, as were monomorphic markers. Markers were characterized for deviation from Hardy-Weinberg proportions in each population using the *hw.test* of pegas, but deviating markers were not removed, only flagged.

The dataset was converted to a genlight object using dartR (v.2.0.4; Gruber *et al.* 2018). Inter-individual relatedness between individuals was estimated using the Ritland statistic (Ritland 1996) as implemented in related (v.1.0; Pew *et al.* 2015). First-degree relatedness values were estimated using known parent-offspring pairs in the VIU dataset (*see below*). Within population pairs with relatedness values above the lower range of this known first-degree relative relatedness statistic were subjected to a process of elimination where one individual per pair was removed until no pairs remained above the threshold. The purged first-degree relative dataset was then analyzed by principal components analysis (PCA) using the function *glPca* in adegenet, retaining PC1-4, and population averaged F_{ST} (Weir & Cockerham 1984) was calculated for any population with more than 15 retained individuals using hierfstat (v.0.5-11; Goudet & Jombart 2022). The full dataset with putative first-degree relatives included was used to identify private alleles in each genetic grouping using poppr (v.2.9.3; Kamvar *et al.* 2014), pooling JPN, FRA, and PEN, and separately pooling the three VIU generations. Different cultured collections were kept as separate groupings in this analysis. The full dataset was also used to calculate per locus mean F_{ST} using the *Fst* function of pegas, and per locus heterozygosity using the *summary* function of adegenet.

Parentage

The genind dataset was converted to rubias format (Moran & Anderson 2019) using the function *genepop_to_rubias* of *simple_pop_stats*. Data preparation and analysis for parentage closely followed vignettes of *close-kin mark recapture-sim* (CKMR-sim; Anderson 2023). Using allele frequencies estimated using the parentage analysis dataset, a simulated set of parent-offspring, full-sib, half-sib, and unrelated relationships were created to determine the power for the amplicon panel to resolve different relationship types, and to estimate expected log likelihood values for the different relationship categories. The dataset was checked for duplicate samples. Parent-offspring relationships and full-sibling relationships in offspring and in broodstock were evaluated using VIU samples (i.e., VIU_F0 and VIU_F1, as well as VIU_F1 and VIU_F2). Additionally, siblings of true parents were included in the analysis to test the power of the amplicon panel to resolve relationships of closely related broodstock individuals. Following analysis and grouping of individuals into putative families, pedigree information was provided and compared against the genetic parentage analysis relationships. Offspring that failed to assign to genotyped parents were tallied to assess false negatives, and incorrect assignments were tallied as false positives; both false negatives and false positives were evaluated in relation to parent-offspring genotyping rates. The number of putative parents was not capped at two, but rather all assignments over the log likelihood cutoff were retained for this analysis. The percentage of the total possible assignments that could have been made was calculated for each family in both parentage analyses. Suspected pedigree or handling errors were inspected alongside genetic sibship analyses and breeding records to reconcile conflicting relationship information. An additional analysis was conducted grouping VIU_F0 and VIU_F1 and comparing against VIU_F2 to evaluate the effect of including a grandparent generation in the analysis. All relationships (FS or PO) above log likelihood score of 5 were retained.

Results

Marker selection, design, and chromosomal positions

The input double digest restriction-site associated DNA sequencing (ddRAD-seq) dataset, containing 366 Pacific oyster samples genotyped at 16,492 SNP loci (single SNP retained per RAD-tag), was pre-filtered for missing data (i.e., present in ≥70% of individuals in all 15 populations), minor allele frequency (MAF) ≥ 0.01, and Hardy-Weinberg equilibrium (Sutherland *et al.* 2020). The dataset was then limited to naturalized British Columbia (BC) samples only, retaining 122 individuals from four populations. A MAF filter was applied to this reduced dataset, retaining 13,438 variants (MAF ≥ 0.01). Markers were mainly selected based on elevated heterozygosity (n = 300; H_{OBS} = 0.41-0.5) and elevated differentiation (n = 300; F_{ST} = 0.046-0.178; Figure S1; Additional File S2). Of the 611 submitted markers, 19 were excluded during

primer design, and 592 were designed into the panel (*Cgig_v.1.0*; see Additional File S3 for marker names, contig identifiers, variant positions and alleles).

Aligning the markers and 200 bp flanking sequences against a chromosome-level assembly of Pacific oyster (GCF_902806645.1; Peñaloza *et al.* 2021) identified 499 markers (i.e., 84.3%) that aligned to a single location in the genome, and this indicated an average (± s.d.) number of markers per chromosome of 49.9 ± 26.3 (range: 15-83; Figure 1). Several significant gaps in chromosome coverage with single mapping markers were observed, and some of these gaps have multiple mapping markers aligned within them (Figure 1).

Panel genotyping, repeatability, and filtering

A pilot study to evaluate the panel comprised 543 samples including technical replicates, or 371 unique samples and 32 negative controls. Samples were genotyped at 592 amplicons across two sequencing chips. Any amplicons that were completely uncalled in either chip were removed from the analysis (n = 7), which left 585 genotyped amplicons. The median of per sample mean marker coverage for the experimental samples in the two chips was 172.5x and 180.1x (n = 371 and 172, respectively), and for the negative control wells was 0.90x and 5.3x. On the second chip, three of the 19 negative control wells had mean marker coverage greater than 40x, suggesting contamination in these wells.

Technical replicate sample pairs (n = 68 pairs) showed average (± s.d.) genotype concordance of 97.5 ± 4.2% for markers genotyped in both samples (mean = 460 markers/pair; total = 31,303 genotypes). Discordant genotypes included homozygous reference - homozygous alternate (n = 4; 0.01%), homozygous reference - heterozygous (n = 153; 0.49%), and homozygous alternate – heterozygous (n = 552; 1.76%). Ten replicate sample pairs had more than 5% of the markers discordantly called; on average, these pairs had 359.5 loci typed in both samples, whereas the remainder pairs on average had 477.7 loci. One replicate pair had genotype discordance above 10%, and this pair was only genotyped at 342 loci. Therefore, samples with lower genotyping rates also had lower concordance.

The 371 unique individuals comprised 181 oysters from the VIU breeding program, 71 from aquaculture programs (i.e., hatcheries or farms), and 118 from naturalized or wild sources (Table 2). After filtering samples for genotyping rate (GR > 50%), 342 unique individuals were retained (Table 2; Figure S2). Filtering loci for genotyping rate (GR > 50%) removed 71 markers (Figure S3), and 50 additional monomorphic markers were also removed, resulting in the retention of 463 markers. Retained markers showed a bimodal H_{OBS} distribution (Figure 2). The upper H_{OBS} grouping ($H_{OBS} > 0.25$; n = 272) is comprised of 89% of the markers selected for high H_{OBS} that passed filters (n = 241). The lower H_{OBS} grouping ($H_{OBS} < 0.25$; n = 198) comprised 83% of the markers selected for high F_{ST} that passed filters, as

well as all custom markers based on presence as private alleles. Per locus H_{OBS} and F_{ST} values are shown in Additional File S4.

Marker Hardy-Weinberg proportions (HWP) were inspected for wild and naturalized populations JPN, PEN, FRA, and CHN finding 19, 35, 35, and 29 variants of the 463 out of HWP, respectively. This includes 13 markers that were out of HWP in two populations, and three markers in three populations. These markers were not removed from the analysis but are flagged for future reference (see Additional File S5).

Genetic parentage analysis

The panel was tested on two separate Vancouver Island University (VIU) spawning events: oyster family run 1 (OFR1) and oyster family run 5 (OFR5). Parents in the OFR1 cross were from naturalized BC populations and are referred to as F0. Parents in the OFR5 cross were offspring of OFR1, and are referred to as F1, and their offspring are referred to as F2.

The OFR5 spawn event analysis includes 15 crosses involving 18 unique parents (nine dams and nine sires) and 98 offspring following filters (Table 4; note: parent BR22 was not genotyped). Parents each had on average 444 genotyped loci. Although pedigree and handling errors were identified, as described below, the following analyses used data with pedigree errors corrected and offspring with mislabeled family identifiers removed. In total, 156 of 186 possible assignments were made (i.e., 84%). Seven families had more than 90% of the possible assignments made. Full parental assignments were made for 67 offspring (i.e., two-parent assignments or one-parent assignment for the family with the missing parent BR22).

False positive assignments (n = 23) occurred in seven of the 15 crosses. These often occurred in addition to the correct assignments and were primarily assignments to siblings of the sire (n = 10) or dam (n = 4). Nine false positives do not have a clear explanation, including five of the offspring of family 5-16 assigning to an individual from family 1-18 (Table 4). All parent-offspring assignments, and full-sib assignments for both generations with accompanying log likelihood ratios are shown in Additional File S6.

Several pedigree or handling errors were identified and corrected in OFR5. Four handling errors were identified where an offspring had the incorrect family label, and these were identified through a combination of parent-offspring and sibship assignments. These incorrectly labeled individuals were dropped from the analysis. Second, family 5-15 was determined to comprise offspring that were either from the correct parents of 5-15 (now labeled 5-15b), or from the parents of family 5-14 (now labeled 5-15a). Third, a pedigree error was identified where the sire of family 5-2 was BR11 (family 1-14) and not BR2 (family 1-12).

The OFR1 spawn event analysis includes 11 crosses involving 12 unique parents (six dams and six sires) and 40 offspring of family OFR1 following filters (Table 5; note: dams BR39 and BR40 were removed

due to low genotyping rates). Unique parents retained after filters each had on average 467.2 genotyped loci. As above, this analysis uses pedigree error corrected data (*see below*). In total, 50 of the 76 possible assignments were made (i.e., 66%). Of the four families with seven or more offspring, two families had good assignment rates (85.7 - 92.9%), and two had poor rates (33.3%-61.1%), resulting in 12 and seven of the 26 total missed assignments, respectively. Notably, the families with poor rates also had low genotype rates in the sire (n = 343 and 390 genotypes) relative to those that performed well (n = 518 and 527 genotypes). Four families with a single offspring each had full two-parent (or one-parent assignments if one parent was missing).

Several pedigree errors were identified and corrected in OFR1. The dam of family 1-9 and 1-13 was BR41 (PSF3) instead of BR42 (PSF4), and the sire of family 1-10 was BR30 (PLM4) instead of BR31 (PSM1). Additionally, although records indicated that BR4 was from family OFR1-1, the correct family was OFR1-10.

An analysis using both OFR1 and OFR5 parents as the possible parents to F2 individuals (i.e., including parents and grandparents of OFR5 offspring), there were several instances of the grandparent generation samples being misidentified as parents (Additional File S6). In total, 11 of the 188 assignments observed were to the grandparent generation, although this includes offspring with more than two parents assigned. Family 5-7 had three instances of assignments to BR30, the grandparent of this family (i.e., parent of BR4). Similarly, family 5-14 had an assignment to grandparent BR41 (i.e., parent of BR24), and family 5-12 had one assignment to grandparent BR36 (i.e., parent of BR20). The F0 generation individual BR34 received one assignment and BR33 received five assignments for no known reason (i.e., these are not the true grandparents of these families). Notably, BR30, BR33, and BR34 were F0 individuals that passed filters but had the lowest genotyping rates, having 324, 343, and 390 loci genotyped, respectively.

Population genetics

The filtered dataset (n = 342 individuals, 463 loci) was then inspected for highly related pairs of individuals within each population (Figure S4). Outlier pairs with elevated relatedness were observed in most populations. The known parent-offspring pairs for OFR5 were used to calculate average relatedness values for the VIU population (n = 178 parent-offspring pairs; average Ritland statistic = 0.39; range: 0.23-0.58). To remove putative first-degree relatives and avoid family effects in PCA and F_{ST}, within-population pairs with Ritland statistic greater than 0.23 across all populations had one of the individuals of the pair removed from the dataset, resulting in the removal of most VIU, DPB, and GUR samples, leaving 11, 6, and 8 individuals per population, respectively. Populations with fewer than 15 individuals were not used for F_{ST} calculation.

A PCA with putative first-degree relatives removed found general overlap of all populations except

CHN and DPB, which were separated from the rest of the samples across PC1 (3.1% variance explained) and PC2 (2.3% variance explained), respectively (Figure 3). Similarly, the Japan translocation lineage (i.e., JPN, PEN, FRA) had low F_{ST}, with the lower range of the 95% confidence intervals overlapping zero (i.e., no differentiation) and the upper range reaching $F_{ST} = 0.016$. By contrast, CHN had higher differentiation, for example JPN-CHN 95% C.I. $F_{ST} = 0.021$-0.046).

Private alleles were empirically identified in the genotyped dataset (with putative first-degree relatives included) at the regional level (i.e., JPN lineage, VIU, DPB, GUR, CHN). In total, 26 private alleles were observed, including some with high frequency. Three markers were included in the panel because they contained private alleles in DPB (Sutherland *et al.* 2020); although different samples were used here, these alleles remained as private alleles in this dataset, including locus 136856 (16 occurrences), locus 39139 (eight occurrences), and locus 538043 (six occurrences). Similarly, locus 336452, included in the panel as a private allele within GUR in Sutherland et al. (2020) was also private in this dataset (10 occurrences). One high frequency private allele was observed in the VIU samples with 11 observations (locus 590965). However, this marker would have been designed based on presence in other BC samples, and so is likely just a rare allele that has been enriched in the VIU broodstock program. A private allele was also observed for each of CHN and JPN with five observations (see full list in Table S1).

Discussion

Using genetic approaches to increase cultured Pacific oyster resiliency may be done through several approaches. First, maintaining genetic diversity overall retains variation needed for rapid adaptation (Barrett & Schluter 2008) and avoids significant inbreeding depression that can impact growth and survival (Wada & Komaru 1994; Evans *et al.* 2004). Second, artificial selection including family-based selection (i.e., breed families showing most favourable phenotypes as selection targets), marker-assisted selection (i.e., select based on genotypes at specific loci), and genomic selection (i.e., breed based on multilocus genotype associations derived from a training set of phenotyped individuals), can be used to improve phenotypic targets of breeding populations, as discussed by Boudry et al. (2021). For effective implementation, these approaches require or significantly benefit from genetic tools; genomic or marker-based selection cannot be done without genetic tools, and maintaining diversity or separate breeding lines for family-based selection benefits from the ability to empirically measure diversity and confirm family relationships.

Maintaining genetic diversity in Pacific oyster breeding populations is important as rapid reductions in genetic diversity can occur (Hedgecock & Sly 1990), as can small sustained reductions that accumulate over time (Carlsson *et al.* 2006; Xiao *et al.* 2011; Varney & Wilbur 2020). Using genetic markers, the

390 effective population size and other measures of diversity such as allelic richness can be directly monitored
391 in the breeding population (Gurney-Smith *et al.* 2017). Allelic richness may be a more appropriate measure
392 to monitor diversity than observed heterozygosity or per-individual inbreeding coefficients (Yu & Guo 2004;
393 Varney & Wilbur 2020; Sutherland *et al.* 2023), in part because these other metrics can be skewed by
394 divergent line crossing that is utilized within the bivalve aquaculture industry. When reduced diversity is
395 detected, solutions may be needed to replenish diversity (Li *et al.* 2007). Some countries, including Canada,
396 have naturalized populations of conspecifics that can be incorporated within breeding programs (Sutherland
397 *et al.* 2020) to replenish lowered diversity and avoid inbreeding. Having tools to empirically measure
398 genetic diversity is important for breeding programs (Gurney-Smith *et al.* 2017) for a range of industry
399 scales (Boudry *et al.* 2021).

400 Family-based selection or crossing of divergent lines both require maintenance of separate genetic
401 groups, and benefit from confirming crosses and pedigree information to correct errors that may occur
402 during animal husbandry or data recording. This has been a subject of concern for over 15 years (Hedgecock
403 & Davis 2007), and has led to the development of high-resolution melt curve genotyping panels (Zhong *et*
404 *al.* 2013; Sun *et al.* 2015). Pedigree errors have been identified in the present work and elsewhere
405 (Hedgecock & Pan 2021) that would not have been identified without the use of a low-cost, effective
406 genotyping panel, and if or when left unchecked, could slow progress in family-based breeding programs
407 (Hedgecock & Pan 2021). The amplicon panel presented here provides a rapid and lower-cost tool to
408 evaluate the parentage of crosses and correct pedigree errors, thereby increasing the precision of family-
409 based selection protocols.

410 There are many potential sources of error that can disrupt maintenance of separate genetic groups
411 and directional family-based selection. They may occur during spawning, larval rearing, or growth phases,
412 and can be caused by incorrect labeling, contaminating families (i.e., inadvertent mixing), and handling and
413 record keeping errors. Many of these errors disrupt pedigree relationships, and if left uncorrected, can
414 compound over time and impede directional selection progress (Hedgecock & Davis 2007). As presented
415 here, we identified several different sources of error in the three generations of the VIU breeding program:
416 (1) mislabeled single offspring; (2) a full-sib family contaminated with a second full-sib family (i.e., the
417 preceding family); and (3) handling or recording errors for the sire of two families within one spawn event.
418 Importantly, as discussed by Hedgecock and Pan (2021), all these errors could impede selective breeding
419 gains were they not detected, and would increase the mismatch between the genetic breeding values
420 expected and the phenotypic outcomes of trials.

421 The amplicon panel may also bring benefits to aquaculture programs in addition to the items
422 described above. Currently, the application of genomic selection in shellfish aquaculture is limited by high
423 genotyping costs per individual. One framework, described and recently evaluated *in silico* for Pacific

oyster by Delomas et al. (2023), is to use whole-genome resequencing of parents, lower-density amplicon panel genotyping of offspring, and imputation to enable more cost-effective application of modern genomic selection methods. Furthermore, field deployments for phenotyping trials can be conducted in common garden replicates, rather than using separate nets that confound spatial and family effects, by using genetic parentage analysis enabled by lower cost genetic marker panels (Allen *et al.* 2020). Finally, future versions of the panels can be designed to include putatively functional markers for marker-assisted selection, for example for shell colour or body size, traits which can increase survival from thermal stress (Raymond *et al.* 2022), or for field survival likely linked to disease resistance (e.g., OsHV-1; Divilov *et al.* 2023).

Future directions

The panel in its present form performed well on the populations included in this analysis. Future work should evaluate the panel in other populations to evaluate any effects of ascertainment bias to determine the limits of its broader use. In this project, all individuals genotyped originated from populations that were included in marker discovery and comprised oysters from the Japan translocation lineage, China, and several commercial aquaculture hatcheries (Sutherland *et al.* 2020). Notably, the actual individuals used in the pilot study were not included in the marker selection dataset, thereby reducing the effect of ascertainment bias on the test dataset, but those from Japan and China were included in marker discovery, and the commercial Deep Bay and Guernsey samples, although different from those in marker discovery, likely include closely related individuals, due to coincident sampling and family structure in these populations. In yet untested populations, other variants within the amplified sequences may prove more powerful for parentage, and these could be characterized via adjustments to the data analysis post sequencing, not requiring any physical changes to the current panel.

Selecting alternate target variants may provide improvements to the panel for parentage purposes without changes to the primer suite. This is apparent given that some markers within Cgig_v.1.0 in the pilot study had low heterozygosity and minor allele frequency, including those selected due to high differentiation potential or presence as private alleles. The pilot also indicates potential for gains, given that some of the OFR1 parents typed at 300 to 400 markers did not receive as many assignments as expected relative to those typed at over 400 markers. Furthermore, in technical replicate pairs, low genotyping rate pairs (e.g., 340-360 markers relative to an average of ~475 markers) had the lowest percent concordance, suggesting that poor genotyping rate is not only an issue for missing data, but the data that is present has a higher chance of miscalls. In the parent-offspring analysis that included both grandparents and parents against F2 offspring, there were higher false positive rates in the low genotyping rate samples for both offspring assigning to grandparents, but also offspring assigning to unrelated individuals (e.g., F0 generation BR33 and BR34 with 343 and 390 genotypes, respectively). It is likely that including higher

value markers (i.e., higher MAF) would increase robustness to missing data relative to the present panel. Higher value markers may also resolve secondary assignments to full-sibs of the true parent without a significant increase in false negatives. Increased statistical power to assign genetic relationships may also be realized through a microhaplotype approach (Baetscher *et al.* 2018), which is an active area of analytic focus for the current panel.

Other advances to the panel are possible but would require changes to the primer suite applied and targeted amplicons. The uniformity of coverage across chromosomes could be improved, filling in gaps observed in the current work, and thereby improving potential for application in genomic selection approaches (Delomas *et al.* 2023). As discussed above, when additional variants with putatively functional roles (or close linkage with causal variants) are identified, such as that recently identified for field survival associated with OsHV-1 outbreaks (e.g., Divilov *et al.* 2023), these will likely be valuable to add to the panel in future iterations. Although rigorous documentation and standardization of locus names is required, the malleability of the amplicon panel is a strength, as loci can be added or dropped far easier than chip based genotyping methods.

For broad-scale utility, the amplicon panel yields the lowest cost per genotyped individual and provides a new tool for emerging Pacific oyster genomic selection programs. However, it is important to also consider the ease of which the data can be analyzed. Although the tab-delimited genotype data produced by the genotyping platform may be easier to handle for those with less bioinformatics experience than high volume sequence data, moving from this output to the formats required for parentage or population genetic analysis does require moderate skills in data manipulation, quality control, and data management. With this in mind, we developed *amplitools*, a version-controlled GitHub repository that is designed to make the use of this panel more accessible, including scripts for converting from raw variant calling output to a standard format used by numerous genetic analysis programs.

Some aspects of documentation within breeding programs will support the use of *amplitools*, such as not including third generations when running parent-offspring analyses, given the observation of grandparents miss-assigning as parents. Including parental sex information may also be beneficial for analysis, although this has not yet been implemented in *amplitools*. Other features of *amplitools*, such as preparing amplicons for design purposes, or maintaining a database, are in ongoing development if not yet currently implemented. Additionally, we developed the associated repository *amplitargets* to keep and version control target files for genotyping, as well as any notes regarding changes between versions of SNP target files. Although the primer panel, *amplitools* and *amplitargets* are currently implemented for the AmpliSeq (ThermoFisher Scientific) platform only, given that the source marker sequences and target variants are all provided here, and therefore could be used for future panel development on other platforms, the code could also be adapted and expanded to other platforms as needed.

Conclusions

The Pacific oyster Cgig_v.1.0 amplicon panel is an effective tool for characterizing genetic variation, genetic similarity among populations, and relatedness between individuals, including for genetic parentage analysis applications. Combined with workflows provided here, this tool could benefit small and large-scale breeding programs by allowing the monitoring of genetic variation and for confirming the family identity of offspring. The amplicon panel is a useful, lower-cost genotyping tool that will evolve in future iterations, specifically through the use of targeting novel variants and microhaplotypes *in silico*, as well as more substantial revisions such as including putatively adaptive loci as they are identified, and achieving more equal coverage across the genome by filling any chromosomal gaps.

Acknowledgements

Thanks to Chen Yin Walker (Vancouver Island University), Claire Rycroft (University of British Columbia), and the Molecular Genetics Lab of Fisheries and Oceans Canada for DNA extraction for the study. Thanks to those who collected or provided samples for this project, including Dr. Kristi Miller (DFO), Rob Saunders (RKS Labs), Prof. Nicolas Bierne and Dr. Ismaël Bernard (Eurêka Modélisation), Prof. Li Li and Dr. Sheng Liu, David Howell, J.P. Hastey (Nova Harvest Ltd.), Keith Reid (Odyssey Shellfish), Brian Yip (Fanny Bay Oysters), and Bruce Evans. This work was funded by the Gordon and Betty Moore Foundation (GBMF#5600) to CAS. This research used resources provided by the SCINet project of the USDA Agricultural Research Service, ARS project number 0500-00093-001-00-D. Funding for NFT was provided by USDA ARS CRIS project 2076-63000-005-000-D. Mention of trade names or commercial products in this publication is solely for the purpose of providing specific information and does not imply recommendation or endorsement by the U.S. Department of Agriculture.

Competing Interests

Ben Sutherland is affiliated with Sutherland Bioinformatics. The author has no competing financial interests to declare. Some authors affiliated with ThermoFisher Scientific have potential conflicts considering that the AgriSeq Targeted Genotyping by Sequencing solutions and associated Oligo panels that were designed and validated in the study are offered by ThermoFisher Scientific. However, the selection of markers, and data analysis was primarily conducted by other authors. The authors declare that the research was conducted in a scientific manner without any commercial considerations that could be construed as potential conflict of interest and further declare no other conflicts of interest. The other authors declare that no competing interests exist.

Data Availability

The amplicon marker panel referred to here as Cgig_v.1.0 is available commercially from ThermoFisher Scientific as SKU A58237 AGRISEQ PACIFIC OYSTER PANEL. The input data required for marker selection (single SNP per locus; plink and VCF formats), the VariantCaller multilocus genotypes for the pilot study, alignment results of submitted sequences against the reference genome, and sample population annotation is available on FigShare, DOI: 10.6084/m9.figshare.23646471

Three repositories support this project:

Manuscript code repository: https://github.com/bensutherland/ms_oyster_panel

amplitools: https://github.com/bensutherland/amplitools

amplitargets: https://github.com/bensutherland/amplitargets

Additional Files

Additional File S1. Amplicon panel design file including target variant with the reference and alternate allele and flanking genomic sequence as well as genomic coordinates.

Additional File S2. Marker names from ddRAD-seq study, reason for selection in panel, and panel marker name.

Additional File S3. Target SNP file with Cgig_v.1.0 target sites used for variant calling.

Additional File S4. Per locus H_{OBS} and F_{ST} from the amplicon panel pilot study.

Additional File S5. Per locus Hardy-Weinberg equilibrium statistics for wild or naturalized populations.

Additional File S6. Parent-offspring relationships for F0 vs. F1 and F1 vs. F2, as well as sibship within each generation.

Supplemental Results. Figure S1-S4 and Table S1, including RAD-seq marker statistics, missing data per individual or marker, interindividual relatedness, and private alleles, respectively.

REFERENCES

Allen SK, Rexroad C, Rheault R (2020) Improving production of the eastern oyster *Crassostrea virginica* through coordination of genetic improvement programs, research, and technology transfer activities. *Journal of Shellfish Research* **39**, 175-179, 175.

Anderson EC (2023) CKMRsim. https://github.com/eriqande/CKMRsim.

Arbelaez JD, Dwiyanti MS, Tandayu E, Llantada K, Jarana A, Ignacio JC, Platten JD, Cobb J, Rutkoski JE, Thomson MJ, Kretzschmar T (2019) 1k-RiCA (1K-Rice Custom Amplicon) a novel genotyping amplicon-based SNP assay for genetics and breeding applications in rice. *Rice* **12**, 55.

Baetscher DS, Clemento AJ, Ng TC, Anderson EC, Garza JC (2018) Microhaplotypes provide increased power from short-read DNA sequences for relationship inference. *Molecular Ecology Resources* **18**, 296-305.

Barrett RDH, Schluter D (2008) Adaptation from standing genetic variation. *Trends in Ecology & Evolution* **23**, 38-44.

Barros P, Sobral P, Range P, Chícharo L, Matias D (2013) Effects of sea-water acidification on fertilization and larval development of the oyster *Crassostrea gigas*. *Journal of Experimental Marine Biology and Ecology* **440**, 200-206.

Barton A, Hales B, Waldbusser GG, Langdon C, Feely RA (2012) The Pacific oyster, *Crassostrea gigas*, shows negative correlation to naturally elevated carbon dioxide levels: Implications for near-term ocean acidification effects. *Limnology and Oceanography* **57**, 698-710.

Beacham TD, Wallace C, MacConnachie C, Jonsen K, McIntosh B, Candy JR, Withler RE (2018) Population and individual identification of Chinook salmon in British Columbia through parentage-based tagging and genetic stock identification with single nucleotide polymorphisms. *Canadian Journal of Fisheries and Aquatic Sciences* **75**, 1096-1105.

Botta R, Asche F, Borsum JS, Camp EV (2020) A review of global oyster aquaculture production and consumption. *Marine Policy* **117**, 103952.

Boudry P, Allal F, Aslam ML, Bargelloni L, Bean TP, Brard-Fudulea S, Brieuc MSO, Calboli FCF, Gilbey J, Haffray P, Lamy J-B, Morvezen R, Purcell C, Prodöhl PA, Vandeputte M, Waldbieser GC, Sonesson AK, Houston RD (2021) Current status and potential of genomic selection to improve selective breeding in the main aquaculture species of International Council for the Exploration of the Sea (ICES) member countries. *Aquaculture Reports* **20**, 100700.

Carlsson J, Morrison CL, Reece KS (2006) Wild and aquaculture populations of the eastern oyster compared using microsatellites. *Journal of Heredity* **97**, 595-598.

Cheng BS, Chang AL, Deck A, Ferner MC (2016) Atmospheric rivers and the mass mortality of wild oysters: insight into an extreme future? *Proceedings of the Royal Society B: Biological Sciences* **283**, 20161462.

Csernák E, Molnár J, Tusnády GE, Tóth E (2017) Application of targeted next-generation sequencing, TruSeq custom amplicon assay for molecular pathology diagnostics on formalin-fixed and paraffin-embedded samples. *Applied Immunohistochemistry & Molecular Morphology* **25**, 460-466.

Delomas TA, Hollenbeck CM, Matt JL, Thompson NF (2023) Evaluating cost-effective genotyping strategies for genomic selection in oysters. *Aquaculture* **562**, 738844.

Divilov K, Merz N, Schoolfield B, Green TJ, Langdon C (2023) Marker-assisted selection in a Pacific oyster population for an antiviral QTL conferring increased survival to OsHV-1 mortality events in Tomales Bay. *Aquaculture* **567**, 739291.

Evans F, Matson S, Brake J, Langdon C (2004) The effects of inbreeding on performance traits of adult Pacific oysters (*Crassostrea gigas*). *Aquaculture* **230**, 89-98.

Goudet J, Jombart T (2022) hierfstat: estimation and tests of hierarchical F-statistics.

Green KM, Spalding AK, Ward M, Levine A, Wolters EA, Hamilton SL, Rice L (2023) Oregon shellfish farmers: Perceptions of stressors, adaptive strategies, and policy linkages. *Ocean & Coastal*

Management **234**, 106475.
Green TJ, Siboni N, King WL, Labbate M, Seymour JR, Raftos D (2019) Simulated marine heat wave alters abundance and structure of Vibrio populations associated with the Pacific oyster resulting in a mass mortality event. *Microbial Ecology* **77**, 736-747.
Gruber B, Unmack PJ, Berry OF, Georges A (2018) dartr: An r package to facilitate analysis of SNP data generated from reduced representation genome sequencing. *Molecular Ecology Resources* **18**, 691-699.
Guo X (2009) Use and exchange of genetic resources in molluscan aquaculture. *Reviews in Aquaculture* **1**, 251-259.
Gurney-Smith HJ, Wade AJ, Abbott CL (2017) Species composition and genetic diversity of farmed mussels in British Columbia, Canada. *Aquaculture* **466**, 33-40.
Gutierrez AP, Turner F, Gharbi K, Talbot R, Lowe NR, Peñaloza C, McCullough M, Prodöhl PA, Bean TP, Houston RD (2017) Development of a medium density combined-species SNP array for Pacific and European oysters (*Crassostrea gigas* and *Ostrea edulis*). *G3 Genes|Genomes|Genetics* **7**, 2209-2218.
Gutierrez AP, Bean TP, Hooper C, Stenton CA, Sanders MB, Paley RK, Rastas P, Bryrom M, Matika O, Houston RD (2018) A genome-wide association study for host resistance to Ostreid herpesvirus in Pacific oysters (*Crassostrea gigas*). *G3 Genes|Genomes|Genetics* **8**, 1273-1280.
Hedgecock D, Sly F (1990) Genetic drift and effective population sizes of hatchery-propagated stocks of the Pacific oyster, *Crassostrea gigas*. *Aquaculture* **88**, 21-38.
Hedgecock D (1994) Does variance in reproductive success limit effective population sizes of marine organisms. *Genetics and evolution of aquatic organisms* **122**, 122-134.
Hedgecock D, McGoldrick DJ, Bayne BL (1995) Hybrid vigor in Pacific oysters: an experimental approach using crosses among inbred lines. *Aquaculture* **137**, 285-298.
Hedgecock D, Gaffney PM, Goulletquer P, Guo X, Reece K, Warr GW (2005) The case for sequencing the Pacific oyster genome. *Journal of Shellfish Research* **24**, 429-441, 413.
Hedgecock D, Davis JP (2007) Heterosis for yield and crossbreeding of the Pacific oyster *Crassostrea gigas*. *Aquaculture* **272**, S17-S29.
Hedgecock D, Pudovkin AI (2011) Sweepstakes reproductive success in highly fecund marine fish and shellfish: a review and commentary. *Bulletin of Marine Science* **87**, 971-1002.
Hedgecock D, Pan FTC (2021) Genetic divergence of selected and wild populations of Pacific oysters (*Crassostrea gigas*) on the West Coast of North America. *Aquaculture* **530**, 735737.
Jombart T, Ahmed I (2011) adegenet 1.3-1: new tools for the analysis of genome-wide SNP data. *Bioinformatics* **27**, 3070-3071.
Kamvar ZN, Tabima JF, Grünwald NJ (2014) poppr: an R package for genetic analysis of populations with clonal, partially clonal, and/or sexual reproduction. *PeerJ* **2**, e281.
King WL, Jenkins C, Seymour JR, Labbate M (2019) Oyster disease in a changing environment: decrypting the link between pathogen, microbiome and environment. *Marine Environmental Research* **143**, 124-140.
Langmead B, Salzberg SL (2012) Fast gapped-read alignment with Bowtie 2. *Nature Methods* **9**, 357-359.
Lapègue S, Harrang E, Heurtebise S, Flahauw E, Donnadieu C, Gayral P, Ballenghien M, Genestout L, Barbotte L, Mahla R, Haffray P, Klopp C (2014) Development of SNP-genotyping arrays in two shellfish species. *Molecular Ecology Resources* **14**, 820-830.
Li C, Wang J, Song K, Meng J, Xu F, Li L, Zhang G (2018) Construction of a high-density genetic map and fine QTL mapping for growth and nutritional traits of *Crassostrea gigas*. *BMC Genomics* **19**, 626.
Li Q, Xu K, Yu R (2007) Genetic variation in Chinese hatchery populations of the Japanese scallop (*Patinopecten yessoensis*) inferred from microsatellite data. *Aquaculture* **269**, 211-219.
Martínez-García MF, Ruesink JL, Grijalva-Chon JM, Lodeiros C, Arreola-Lizárraga JA, de la Re-Vega E, Varela-Romero A, Chávez-Villalba J (2022) Socioecological factors related to aquaculture introductions and production of Pacific oysters (*Crassostrea gigas*) worldwide. *Reviews in*

Aquaculture **14**, 613-629.
Meek MH, Larson WA (2019) The future is now: Amplicon sequencing and sequence capture usher in the conservation genomics era. *Molecular Ecology Resources* **19**, 795-803.
Moran BM, Anderson EC (2019) Bayesian inference from the conditional genetic stock identification model. *Canadian Journal of Fisheries and Aquatic Sciences* **76**, 551-560.
Paradis E (2010) pegas: an R package for population genetics with an integrated–modular approach. *Bioinformatics* **26**, 419-420.
Peñaloza C, Gutierrez AP, Eöry L, Wang S, Guo X, Archibald AL, Bean TP, Houston RD (2021) A chromosome-level genome assembly for the Pacific oyster *Crassostrea gigas*. *GigaScience* **10**.
Pew J, Muir PH, Wang J, Frasier TR (2015) related: an R package for analysing pairwise relatedness from codominant molecular markers. *Mol Ecol Resour* **15**, 557-561.
Plough LV (2016) Genetic load in marine animals: a review. *Current Zoology* **62**, 567-579.
Qi H, Song K, Li C, Wang W, Li B, Li L, Zhang G (2017) Construction and evaluation of a high-density SNP array for the Pacific oyster (*Crassostrea gigas*). *PLOS ONE* **12**, e0174007.
Qi H, Li L, Zhang G (2021) Construction of a chromosome-level genome and variation map for the Pacific oyster *Crassostrea gigas*. *Molecular Ecology Resources* **21**, 1670-1685.
Quinlan AR, Hall IM (2010) BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841-842.
R Core Team (2023) R: A Language and Environment for Statistical Computing, Vienna, Austria.
Raymond WW, Barber JS, Dethier MN, Hayford HA, Harley CDG, King TL, Paul B, Speck CA, Tobin ED, Raymond AET, McDonald PS (2022) Assessment of the impacts of an unprecedented heatwave on intertidal shellfish of the Salish Sea. *Ecology* **103**, e3798.
Ritland K (1996) Estimators for pairwise relatedness and individual inbreeding coefficients. *Genetics Research* **67**, 175-185.
Salvi D, Mariottini P (2017) Molecular taxonomy in 2D: a novel ITS2 rRNA sequence-structure approach guides the description of the oysters' subfamily Saccostreinae and the genus Magallana (Bivalvia: Ostreidae). *Zoological Journal of the Linnean Society* **179**, 263-276.
Sauvage C, Bierne N, Lapègue S, Boudry P (2007) Single nucleotide polymorphisms and their relationship to codon usage bias in the Pacific oyster *Crassostrea gigas*. *Gene* **406**, 13-22.
Sun X, Shin G, Hedgecock D (2015) Inheritance of high-resolution melting profiles in assays targeting single nucleotide polymorphisms in protein-coding sequences of the Pacific oyster *Crassostrea gigas*: Implications for parentage assignment of experimental and commercial broodstocks. *Aquaculture* **437**, 127-139.
Sutherland BJG, Rycroft C, Ferchaud AL, Saunders R, Li L, Liu S, Chan AM, Otto SP, Suttle CA, Miller KM (2020) Relative genomic impacts of translocation history, hatchery practices, and farm selection in Pacific oyster *Crassostrea gigas* throughout the Northern Hemisphere. *Evolutionary Applications* **13**, 1380-1399.
Sutherland BJG, Itoh N, Gilchrist K, Boyle B, Roth M, Green TJ (2023) Genomic diversity of wild and cultured Yesso scallop *Mizuhopecten yessoensis* from Japan and Canada. *bioRxiv*, 2023.2005.2031.540437.
Varney RL, Wilbur AE (2020) Analysis of genetic variation and inbreeding among three lines of hatchery-reared *Crassostrea virginica* broodstock. *Aquaculture* **527**, 735452.
Wada KT, Komaru A (1994) Effect of selection for shell coloration on growth rate and mortality in the Japanese pearl oyster, *Pinctada fucata martensii*. *Aquaculture* **125**, 59-65.
Watson S-A, Southgate PC, Tyler PA, Peck LS (2009) Early larval development of the Sydney rock oyster *Saccostrea glomerata* under near-future predictions of CO_2-driven ocean acidification. *Journal of Shellfish Research* **28**, 431-437, 437.
Weir BS, Cockerham CC (1984) Estimating F-statistics for the analysis of population structure. *Evolution* **38**, 1358-1370.
Wickham H (2016) *ggplot2: elegant graphics for data analysis* Springer-Verlag New York.
Xiao J, Cordes JF, Moss JA, Reece KS (2011) Genetic diversity in U.S. hatchery stocks of *Crassostrea*

ariakensis (Fujita, 1913) and comparison with natural populations in Asia. *Journal of Shellfish Research* **30**, 751-760, 710.

Yu Z, Guo X (2004) Genetic analysis of selected strains of eastern oyster (*Crassostrea virginica* Gmelin) using AFLP and microsatellite markers. *Marine Biotechnology* **6**, 575-586.

Zhang G, Fang X, Guo X, Li L, Luo R, Xu F, Yang P, Zhang L, Wang X, Qi H, Xiong Z, Que H, Xie Y, Holland PWH, Paps J, Zhu Y, Wu F, Chen Y, Wang J, Peng C, Meng J, Yang L, Liu J, Wen B, Zhang N, Huang Z, Zhu Q, Feng Y, Mount A, Hedgecock D, Xu Z, Liu Y, Domazet-Lošo T, Du Y, Sun X, Zhang S, Liu B, Cheng P, Jiang X, Li J, Fan D, Wang W, Fu W, Wang T, Wang B, Zhang J, Peng Z, Li Y, Li N, Wang J, Chen M, He Y, Tan F, Song X, Zheng Q, Huang R, Yang H, Du X, Chen L, Yang M, Gaffney PM, Wang S, Luo L, She Z, Ming Y, Huang W, Zhang S, Huang B, Zhang Y, Qu T, Ni P, Miao G, Wang J, Wang Q, Steinberg CEW, Wang H, Li N, Qian L, Zhang G, Li Y, Yang H, Liu X, Wang J, Yin Y, Wang J (2012) The oyster genome reveals stress adaptation and complexity of shell formation. *Nature* **490**, 49-54.

Zhong X, Li Q, Yu H, Kong L (2013) Development and validation of single-nucleotide polymorphism markers in the Pacific oyster, *Crassostrea gigas*, using High-resolution Melting analysis. *Journal of the World Aquaculture Society* **44**, 455-465.

FIGURES

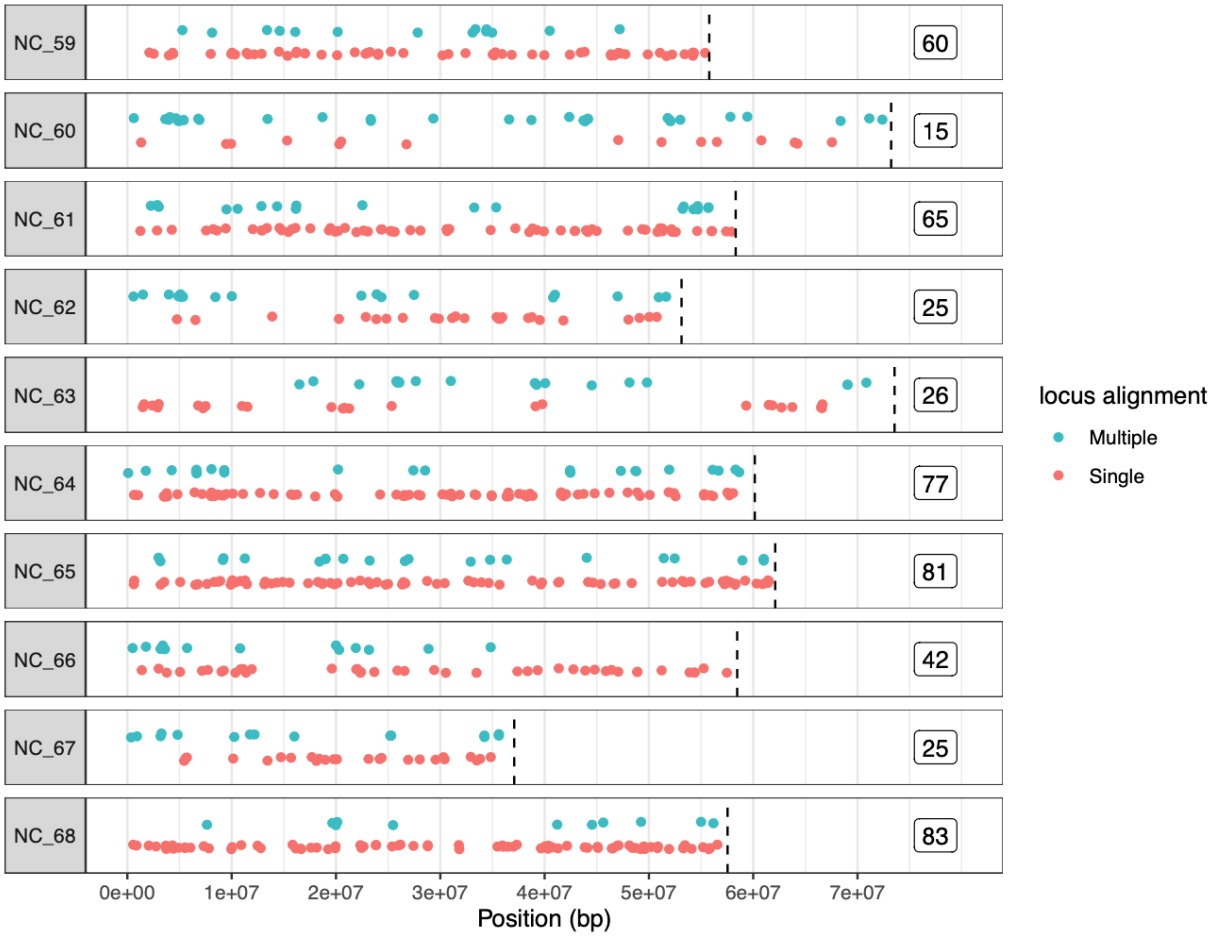

Figure 1. The marker panel design sequences were aligned against a chromosome reference genome to observe coverage across the ten chromosomes (GCF_902806645.1; Peñaloza *et al.* 2021). Although most markers mapped to a single position in the genome (n = 499; red/lower), there were 55 multi-mapping amplicons (blue/upper) that aligned to more than one position in the genome. Ends of chromosomes are shown by vertical hatched lines, and NCBI chromosome names abbreviations are given in the grey boxes to the left of the plot (e.g., NC_047559.1 to NC_59).

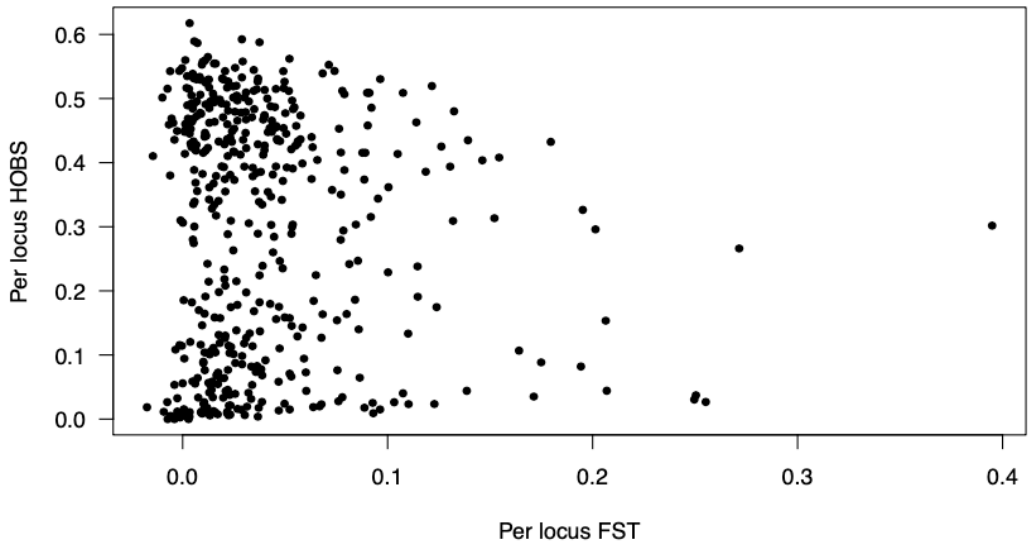

Figure 2. Per locus observed heterozygosity (H_{OBS}) plotted alongside genetic differentiation (F_{ST}) based on data from the pilot study dataset (technical replicate samples removed, filtered genotypes).

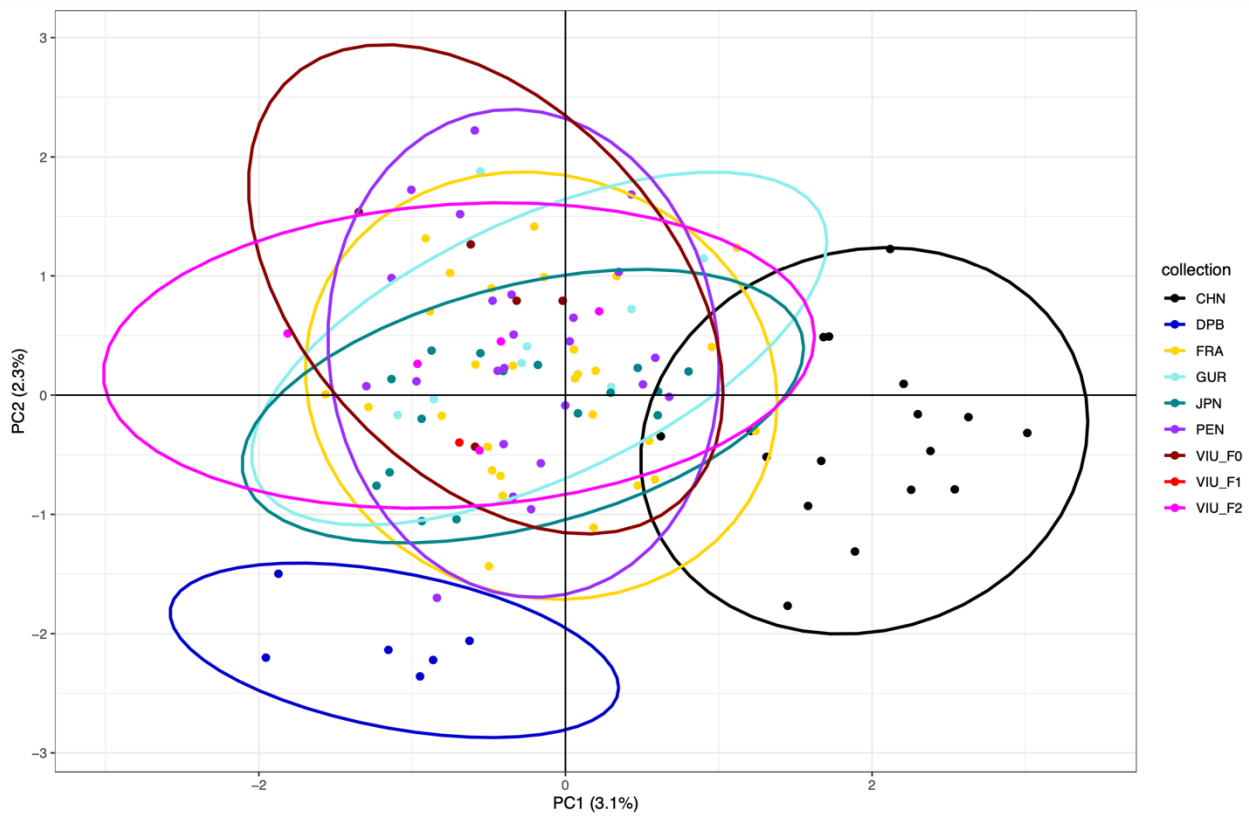

Figure 3. PCA of pilot study samples clustered by filtered genotypes after putative first degree relatives were removed. Population abbreviations: VIU = Vancouver Island University (hatchery); DPB = Deep Bay (farm); PEN = Pendrell (naturalized); GUR = Guernsey (hatchery); FRA = France (naturalized); JPN = Japan (wild); CHN = China (wild).

TABLES

Table 1. Markers in the Cgig_v.1.0 amplicon panel, including those submitted for design, those that were designed, and those that passed quality control including minor allele frequency filters in the pilot dataset.

Selection criterion	Markers (n)	Designed (n)	Quality controlled in test dataset
F_{ST}	291	278	180
Heterozygosity	291	285	263
F_{ST} *and heterozygosity*	9	9	7
Private alleles, DPB	15	15	10
Private alleles, GUR	5	5	3
Total	**611**	**592**	**463**

Table 2. Pilot study samples genotyped and passing filters for each collection, and the number of polymorphic variants per collection. Note: all Canadian collections were taken in British Columbia

Collection	Type	Country of origin	Country collected	Genotyped (n)	Filtered (n)	Retained (%)
Pendrell (PEN)	Naturalized	Canada	Canada	31	30	97%
France (FRA)	Naturalized	France	France	32	32	100%
Japan (JPN)	Wild	Japan	Japan	22	22	100%
China (CHN)	Wild	China	China	33	33	100%
China (QDC)	Commercial hatchery	China	China	8	8	100%
Guernsey (GUR)	Farm / commercial hatchery	United Kingdom	Canada	32	20	63%
Deep Bay (DPB)	Farm / commercial hatchery	*unknown*	Canada	32	32	100%
VIU F0	University hatchery	*unknown*	Canada	16	14	88%
VIU F1	University hatchery	*unknown*	Canada	53	49	92%
VIU F2	University hatchery	*unknown*	Canada	112	102	91%
Totals				**371**	**342**	

Table 3. Genetic differentiation (F_{ST}) population averages shown as lower (bottom left) and upper (top right) 95% confidence limits. Estimates from amplicon panel data using 1000 bootstraps. PEN = Pendrell (naturalized); FRA = France (naturalized); JPN = Japan (wild); CHN = China (wild).

	PEN	**FRA**	**JPN**	**CHN**
PEN	-	0.0127	0.0155	0.0451
FRA	0.0010	-	0.0045	0.0357
JPN	0.0026	-0.0042	-	0.0457
CHN	0.0241	0.0171	0.0205	-

Table 4. Oyster family run 5 (OFR5) parentage analysis showing the dam and sire identifier, family ID (fam.), and number loci genotyped. The number of genotyped offspring and the correct two- or one-parent assignments are shown, along with any comments about the assignments. FP = false positive.

OFR5 family	Dam [fam.] (loci)	Sire [fam.] (loci)	Off-spring	2p assign.	1p assign.	% ident-ified	Incorrect assignments
OFR5_2	BR19 [2-6] (448)	BR11 [1-14] (457)	8	5	3	81.3	
OFR5_3	BR14 [2-10] (418)	BR2 [1-12] (448)	7	3	4	71.4	
OFR5_4	BR14 [2-10] (418)	BR11 [1-14] (457)	7	6	1	92.9	
OFR5_5	BR3 [1-4] (438)	BR10 [1-18] (450)	7	3	4	71.4	FP: 1 x pat. sib
OFR5_7	BR17 [2-11] (455)	BR4 [1-10] (442)	8	7	1	93.8	
OFR5_8	BR3 [1-4] (438)	BR7 [2-11] (438)	8	5	3	81.3	
OFR5_9	BR9 [1-16] (434)	BR23 [2-6] (450)	7	6	1	92.9	
OFR5_10	BR25 [1-19] (446)	BR23 [2-6] (450)	8	5	3	71.4	FP: 4 x mat. sib, 1 x BR9
OFR5_12	BR14 [2-10] (418)	BR20 [1-19] (450)	7	7	0	100.0	FP: 3 x pat. sib
OFR5_13	BR22 [2-14] (0)	BR26 [1-9] (455)	5	0	5	100.0	
OFR5_14	BR24 [1-13] (439)	BR2 [1-12] (448)	7	3	4	71.4	FP: 3 x pat. sib, 1 x BR26, 1 x BR21
OFR5_15a	BR24 [1-13] (439)	BR2 [1-12] (448)	4	4	0	100.0	FP: 2 x pat. sib, 1 x BR11
OFR5_15b	BR18 [2-13] (448)	BR10 [1-18] (450)	3	3	0	100.0	FP: 1 x pat. sib
OFR5_16	BR22 [2-14] (0)	BR2 [1-12] (448)	5	0	2	40.0	FP: 5 x [1-18]
OFR5_17	BR21 [1-1] (441)	BR5 [2-21] (443)	7	4	3	78.6	

753 **Table 5.** Oyster family run 1 (OFR1) parentage analysis showing the dam and sire identifier, family ID
754 (fam), and number loci genotyped. The number of genotyped offspring and the correct two- or one-parent
755 assignments are shown, along with any comments about the assignments. FP = false positive. * denotes that
756 one parent is missing, so a one-parent assignment is all that is possible.

OFR1 Family	Dam [fam.] (loci)	Sire [fam.] (loci)	Off-spring	2p assign.	1p assign.	% ID	Incorrect assignments
OFR1_1	BR39 [PSF1] (0)	BR27 [PLM1] (546)	1	0	1*	100.0	
OFR1_10	BR41 [PSF3] (509)	BR30 [PLM4] (324)	1	0	1	50.0	
OFR1_12	BR35 [PLF1] (515)	BR31 [PSM1] (518)	7	5	2	85.7	FP: 1 x BR33
OFR1_13	BR41 [PSF3] (509)	BR32 [PSM2] (527)	1	1	0	100.0	FP: 1 x BR33
OFR1_14	BR35 [PLF1] (515)	BR32 [PSM2] (527)	7	6	1	92.9	
OFR1_16	BR35 [PLF1] (515)	BR33 [PSM3] (343)	1	0	1	50.0	Low target geno
OFR1_18	BR37 [PLF3] (514)	BR33 [PSM3] (343)	9	0	6	33.3	Low target geno
OFR1_19	BR36 [PLF2] (486)	BR34 [PSM4] (390)	9	2	7	61.1	Low target geno
OFR1_4	BR39 [PSF1] (0)	BR33 [PLM3] (343)	2	0	0*	0.0	Low target geno
OFR1_7	BR40 [PSF2] (0)	BR30 [PLM4] (324)	1	0	1*	100.0	
OFR1_9	BR41 [PSF3] (509)	BR30 [PLM4] (324)	1	1	0	100.0	

757