Genomic diversity of wild and cultured Yesso scallop *Mizuhopecten yessoensis* from Japan and Canada

Ben J. G. Sutherland[1,2], Naoki Itoh[3], Korrina Gilchrist[1], Brian Boyle[4], Myron Roth[5], Timothy J. Green[1]

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

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

[3] Laboratory of Fish Diseases, Graduate School of Agricultural and Life Sciences, The University of Tokyo, 1-1-1, Yayoi, Bunkyo, Tokyo 113-8657, Japan

[4] Plateforme d'Analyses Génomiques, Institut de Biologie Intégrative et des Systèmes (IBIS), Université Laval, Québec, QC, Canada

[5] BC Ministry of Agriculture and Food, P.O. Box 9120, Victoria, BC V8W 9B4, Canada

Author for correspondence: BJGS
Sutherland Bioinformatics, Lantzville, BC, Canada V0R 2H0
Email: Sutherland.Bioinformatics@protonmail.com

Data deposition: Raw sequencing data has been uploaded to SRA under BioProject PRJNA947158, BioSamples SAMN33843243-SAMN33843431.

Abstract

The Yesso scallop *Mizuhopecten yessoensis* is an important aquaculture species that was introduced to Western Canada from Japan to establish an economically viable scallop farming industry. This highly fecund species has been propagated in Canadian aquaculture hatcheries for the past 40 years, raising questions about genetic diversity and genetic differences among hatchery stocks. In this study, we compare cultured Canadian and wild Japanese populations of Yesso scallop using double-digest restriction site-associated DNA (ddRAD)-sequencing to genotype 21,048 variants in 71 wild-caught scallops from Japan, 65 scallops from the Vancouver Island University breeding population, and 37 scallops obtained from a commercial farm off Vancouver Island, British Columbia. The wild scallops are largely comprised of equally unrelated individuals, whereas cultured scallops are comprised of multiple families of related individuals. The polymorphism rate estimated in wild scallops was 1.7%, whereas in the cultured strains it varied between 1.35% and 1.07%. Interestingly, heterozygosity rates were highest in the cultured populations, which is likely due to shellfish hatchery practices of crossing divergent strains to gain benefits of heterosis and to avoid inbreeding. Evidence of founder effects and drift were observed in the cultured strains, including high genetic differentiation between cultured populations and between cultured populations and the wild population. Cultured populations had effective population sizes ranging from 9-26 individuals whereas the wild population was estimated at 25-50K individuals. Further, a depletion of low frequency variants was observed in the cultured populations. These results indicate significant genetic diversity losses in cultured scallops in Canadian breeding programs.

Keywords: aquaculture; bivalve; genetic diversity; *Mizuhopecten*; RAD-sequencing; scallop

Introduction

Marine bivalves (Class Bivalvia) have high genetic diversity (Plough 2016) both in terms of polymorphism rate (Hedgecock *et al.* 2005) and observed heterozygosity (Solé-Cava & Thorpe 1991). Although this may be in part due to large population sizes and wide planktonic dispersal, it is likely also related to high mutation rates that occur due to the number of meiotic events required to produce millions of eggs per individual (i.e., the Elm-Oyster model; Williams 1975; Plough 2016). There may also be an involvement of transposable element activity related to the elevated polymorphism rate (Zhang *et al.* 2012). Mutations are not effectively purged from the population even if mildly deleterious due to their introduction into the population at high frequency and the low effective population size relative to census size from sweepstakes reproductive success (SRS; Hedgecock 1994; Hedgecock & Pudovkin 2011; Plough 2016). This genetic load explains the severe consequences of inbreeding depression in cultured bivalves, as well as the striking heterosis observed when counteracting it through outbreeding (Hedgecock *et al.* 1995; Hedgecock & Davis 2007).

The Yesso scallop *Mizuhopecten yessoensis* is an economically important aquaculture species and a valuable commercial protein source. Native to Japan, it has been grown in culture in British Columbia (BC; Canada) and China since the 1980s (Bourne 2000; Li *et al.* 2007). In China it is currently produced at over 100,000 tonnes per year, but BC has not reached its production potential, despite favourable environmental conditions (Holden *et al.* 2019). BC production of Yesso scallop averaged 89 tonnes per annum in 2017-2020, but decreased to 52 tonnes in 2021 (DFO 2023). Shellfish aquaculture, including scallop production, has realized benefit and strong potential benefit to remote coastal communities economically in terms of jobs and commercial value, although expansion will require local logistical considerations as well as diverse stakeholder engagement (Holden *et al.* 2019). Yesso scallops take two to three years to reach market size in BC (Bower *et al.* 1999), and high mortality rates challenge the industry, limiting investment, and thus negatively impacting large and small growers, including First Nations-led companies. High mortality occurs in the hatchery, but also at later grow-out stages, impacting larger scallops that have already received significant production effort. On the farm, mortality may be caused by predation by the flatworm *Pseudostylochus orientalis*, and infection with the bacteria *Francisella halioticida* or the protozoan *Perkinsus qugwadi* (Bower *et al.* 1999; Bourne 2000; Meyer *et al.* 2017). The industry is therefore now looking towards selective breeding as well as ensuring fitness of broodstock lines to reduce mortalities and improve yields.

Loss of genetic diversity in cultured lineages may occur by founder effects, drift, and selective breeding. Relatively rapid reductions in genetic diversity are commonly observed in bivalve hatchery lineages (Hedgecock & Sly 1990; Evans *et al.* 2004; Gurney-Smith *et al.* 2017). For example, hatchery lineages of Asian Suminoe oyster *C. ariakensis* were all found to be significantly lower in diversity than

wild populations (Xiao *et al.* 2011); decreases were most pronounced in established lineages (60% decrease in allelic richness compared to wild) relative to more recently established lineages (17-30% decrease). This is also an issue for the cultured eastern oyster *Crassostrea virginica* (Carlsson *et al.* 2006; Varney & Wilbur 2020). In the aforementioned cultured populations of Yesso scallop in China, reduced polymorphism rate and heterozygosity has occurred, which is a major concern for the industry (Li *et al.* 2007). This is particularly problematic when hatchery populations are not able to be replenished with naturalized or wild populations, such as is the case when growing outside of native ranges (e.g., Canada, China). It is an important goal for the industry to retain adaptive potential, and monitor selected strains for diversity loss (Carlsson *et al.* 2006).

Inbreeding depression may result due to a loss of genetic diversity of the cultured lineages, and in bivalves this has been associated with declines in survival and growth rates and increases in deformities. Inbreeding depression reduces survival in the Japanese pearl oyster *Pinctada fucata martensii* (Wada & Komaru 1994), Pacific oyster *Crassostrea gigas* (Evans *et al.* 2004), and Pacific abalone *Haliotis discus hannai* (Kobayashi & Kijima 2010). In Yesso scallop, efforts have been undertaken to explore the genetic mechanisms underlying inbreeding depression through analyzing transcriptome profiles in inbred animals (Zhao *et al.* 2019). Therefore, although selective breeding shows strong potential in Yesso scallop due to ease of controlling the biological cycle in culture, sexual dimorphism, high fecundity, and high standing genetic variation, and has been demonstrated for shell colouration (Zhao *et al.* 2017), efforts must also be put towards avoiding inbreeding depression that can occur through the selective breeding process. The genome is relatively large and complex in the scallop (Family Pectinidae), due to tandem gene duplication and gene family expansions (Kenny *et al.* 2020). However, a high-quality chromosome-level assembly has been constructed for the Yesso scallop (Wang *et al.* 2017), enabling many genetic and genomic tools and approaches to support selective breeding and broodstock monitoring.

To support the Yesso scallop industry in BC, it is first necessary to determine the existing amount of genetic variation contained within cultured lineages. Following the introduction in the 1980s (Bourne 2000), the extent of drift along with the reduced diversity from founder effects for cultured populations will be informative to understand the diversity currently present and therefore the potential for long-term sustainability of the broodstock lineages and adaptive potential during selective breeding. Little or no formal broodstock management plans for the industry have been implemented on a broad scale in BC, and assessing the current state of diversity available will help lead towards such a goal. Here we use double-digest restriction site-associated DNA (ddRAD)-sequencing to characterize scallops from the Vancouver Island University (VIU) breeding program, as well as those commercially obtained from a BC farm, and compare these to wild Yesso scallops obtained from Japan through commercial harvest or research surveys. Comparisons between these populations will indicate the current state of the cultured lineages in BC.

121 Furthermore, these results will identify highly heterozygous markers in both BC and Japan that can then be
122 developed into rapid and low-cost genotyping platforms for parentage and other purposes to support
123 broodstock management and selective breeding in the Yesso scallop.
124

125 2 | Methods

126 2.1 | Sample collections

127 *Mizuhopecten yessoensis* samples were obtained from four different sources (Table 1). Wild scallops from
128 Japan were obtained from marine surveys in Southern Hokkaido in late 2017 through early 2021 (see
129 Additional File S1), and from a commercial source as harvested wild scallops from Northern Hokkaido in
130 2021. BC farmed scallops were also obtained from a commercial provider that were grown near central
131 Vancouver Island, BC. Samples from the Vancouver Island University (VIU) Centre for Shellfish Research
132 breeding program were obtained from animals in late 2022.
133

134 2.2 | DNA extraction, quantification, and quality control

135 Marine survey samples were obtained as pelleted DNA in 100% ethanol and were stored at -80ºC. Samples
136 were purified by centrifugation (13,000 x g for 30 min at 4ºC), followed by two washes with 0.5 ml ice cold
137 75% ethanol, followed by resuspending the dried DNA pellet in molecular grade water. Commercial and
138 breeding program samples were extracted from frozen tissue using the Monarch Genomic DNA Purification
139 Kit (NEB), as per manufacturer's instructions using the enzymatic cleanup protocol with both proteinase K
140 and RNase A, and eluting into 50 µl molecular grade water.
141 All samples were quantified by Qubit dsDNA-BR measured by BioSpectrometer (Eppendorf) with
142 a reusable µCuvette (G.1.0; Eppendorf). Any sample under 20 ng/µl was concentrated using the Genomic
143 DNA Concentrator Columns (Zymo Research), eluted in 12 µl water, and quantified by Qubit. Quality was
144 inspected in representative samples from each collection by 1% agarose gel electrophoresis. Samples were
145 normalized to 20 ng/µl in a 10 µl volume, randomized in 96-well microtitre plates and submitted to the
146 Institute de Biologie Intégrative et des Systèmes (IBIS) at Université Laval for library preparation. Each
147 plate contained a negative control well. Five samples with concentrations between 17.7 – 20 ng/µl were
148 also included in a 10 µl volume (i.e., JPN samples 018, 031, 039, 051, and 082).
149

150 2.3 | Double-digest RAD-sequencing library preparation

151 Selection of restriction enzymes and estimation of required sequencing depth was conducted using
152 SimRAD (Lepais & Weir 2014) using the Yesso scallop reference genome (GCA_002113885.2; Wang *et*
153 *al.* 2017). The estimated haploid genome size for this species is 1.65 Gbp (Anisimova 2007; Gregory 2022),

154 but the total sequence length of the assembly is 0.99 Gbp, suggesting that the assembly represents 59.8%
155 of the expected genome size. This value was used as an expansion factor for the true expected number of
156 fragments estimated by SimRAD. Enzyme options tested were *Pst*I and *Msp*I, or *Nsi*I and *Msp*I, size-
157 selected to retain fragments from 100-250 bp. Digestions with *Nsi*I and *Msp*I were predicted *in silico* to
158 generate 118,850 fragments in the 1.65 Gbp genome, whereas *Pst*I and *Msp*I were predicted to have 50,836
159 fragments; *Pst*I and *Msp*I were selected for digestion.

160 Samples were multiplexed to 32 individuals per chip, with six sequencing chips used. Library
161 preparation was conducted using the semiconductor platform-adapted ddRAD-seq approach (Mascher *et*
162 *al.* 2013) with the additional barcodes and size selection step (Abed *et al.* 2019), as previously described
163 (Sutherland *et al.* 2020), but with alternate enzymes. Multiplexed and size-selected libraries were then
164 sequenced on an Ion Torrent Proton at IBIS using the Ion PI Chip Kit v3 chip (Thermo Fisher), as per
165 manufacturer's instructions. Base calling was conducted with the Torrent Suite software (Thermo Fisher),
166 and multiplexed fastq output files were exported using the *FileExporter* plugin.
167

168 **2.4 | Sequence data preprocessing, alignment, and genotyping**

169 Sequence data processing and genotyping largely followed the *stacks_workflow* repository (E.
170 Normandeau), and all instructions for the analysis is provided (*ms_scallop_popgen*; see *Data Availability*).
171 Sample barcodes and populations were designated in the metadata file (see *Data Availability*). Raw
172 sequence data was quality checked using FastQC (v.0.11.4; Andrews 2010) and MultiQC (v.1.14; Ewels *et*
173 *al.* 2016). Sequences were trimmed using cutadapt (v.1.9.1; Martin 2011) to remove too short reads and
174 adapters, and inspected again with FastQC and MultiQC. The metadata file was used to demultiplex using
175 the *process_radtags* module of Stacks v2 with flags pstI and mspI, and truncating all reads to 80 bp (v.2.62;
176 Rochette *et al.* 2019) in parallel (Tange 2022).

177 Demultiplexed samples were aligned against the reference genome (ASM211388v2; Wang *et al.*
178 2017) using bwa mem (Li 2013) then converted to bam format and sorted using samtools (Danecek *et al.*
179 2021). Read and alignment counts per individual were tallied using custom code (*ms_scallop_popgen*; see
180 *Data Availability*). Samples with fewer than 500,000 reads were removed including the three negative
181 control samples that had between 338-2,101 reads each. The gstacks module of Stacks was then used with
182 the reference genome to genotype the samples using default settings (marukilow, var_alpha: 0.01, gt_alpha:
183 0.05). Following genotyping, the populations module of Stacks was used to filter, retaining loci genotyped
184 in ≥70% of the individuals per population in all three populations (flags: -r 0.7, -p 3). A global minor allele
185 frequency (MAF) filter retained loci with MAF ≥ 0.01 (flag: --min_maf 0.01). To determine the impact of
186 unequal sample size on the mean nucleotide diversity estimated per population, gstacks and populations
187 were run as above but with a rarefied dataset using 37 samples per population for all three populations.

188 Using the full dataset, the per-individual inbreeding coefficient was calculated using vcftools (--het)
189 (Danecek *et al.* 2011), and any outlier individuals with excessive heterozygosity indicating potential
190 genotyping issues were removed. Once outliers were removed, genotyping and filtering was redone with
191 only the retained individuals, as described above.
192 Multi-locus genotypes were exported in VCF and plink formats, where only one of the variants per
193 RAD-tag was retained using the *--write-single-snp* flag of the populations module to produce the single-
194 SNP per locus dataset. Furthermore, microhaplotype genotypes were exported in RADpainter and genepop
195 formats. Plink data was converted from .ped and .map format to .raw format using plink (flags: --recode A
196 --allow-extra-chr; v.1.90b6.26; Purcell *et al.* 2007; Purcell 2022). The plink output was then used as input
197 for population genetic analysis in R (R Core Team 2023).
198

199 **2.5 | Population genetic analysis**

200 Single-SNP per locus genotypes were read into R using the *read.PLINK* function of adegenet (Jombart &
201 Ahmed 2011) and MAF densities calculated by *glMean*. Data were formatted and converted to genind
202 format using the *df2genind* function of adegenet. Functions applied were used from the *simple_pop_stats*
203 repository (see *Data Availability*). Genotyping rate was calculated per sample using the
204 *percent_missing_by_ind* function of *simple_pop_stats*, and individuals were retained if they had less than
205 30% missing data. Per locus F_{ST} was calculated using pegas (Paradis 2010). Observed heterozygosity (H_{OBS})
206 was calculated using adegenet, and a test of Hardy-Weinberg (HW) equilibrium was conducted per locus
207 using pegas. Loci showing significant deviation from HW proportions ($p \leq 0.01$) in any one of three
208 populations were identified and removed from the dataset. Loci with global $H_{OBS} \geq 0.5$ were also removed.
209 The filtered dataset was used as an input to a principal components analysis (PCA) using the *glPca*
210 function of adegenet. Mean population-level F_{ST} was calculated using 1000 bootstraps to generate 95%
211 confidence limits using the boot.ppfst function of hierfstat (Goudet & Jombart 2022). Private alleles were
212 calculated by private_alleles function of poppr (Kamvar *et al.* 2014) using the full, filtered dataset, or the
213 dataset rarefied to retain an equal number of samples for each population using the minimum sample size
214 for the study. Region-specific private alleles were identified by combining all Canadian samples and
215 comparing to the Japan samples. Per locus H_{OBS} was also calculated for each population specifically to
216 identify loci with high heterozygosity in JPN and VIU populations. Effective population size (N_e) was
217 calculated using NeEstimator v.2.0 (Do *et al.* 2014) using both single-SNP per locus data and
218 microhaplotype data. Inter-individual relatedness was calculated using single-SNP per locus data using
219 *related* (Pew *et al.* 2015), retaining only comparisons within a population. Relatedness was also calculated
220 using microhaplotype data using fineRADstructure (Malinsky *et al.* 2018).
221

3 | Results

3.1. Sequencing, genotyping, and filtering

In total, 558,098,638 single-end reads were generated from 189 Yesso scallops sequenced on six Ion Torrent chips. Each chip produced on average (± s.d.) 93.0 ± 1.4 M reads. After quality trimming, 547,450,718 reads (i.e., 98.1%) were retained, and the mean read duplication rate per sample was 86.6%. The mean GC content per sample was 43%, which is higher than the reference genome (i.e., 36.5%; GCF_002113885.1). After demultiplexing by sample, 503.5 M reads were retained (i.e., 92.1% of trimmed reads), where the per-sample average (± s.d.) number of reads was 2.7 ± 1.2 M (Table 1). Samples with the fewest reads were from the Japan (JPN) marine surveys with an average number of reads per sample of 1.7 M (± 0.6 M), and those with the most were from the British Columbia (BC) farm samples with an average number of reads per sample of 3.5 M (± 0.9 M). One JPN marine survey scallop (JPN_075) was removed due to low coverage, and one VIU sample (VIU_002) was removed due to being an extreme H_{OBS} outlier during an initial run of Stacks (F = -0.1938).

Aligning samples against the reference genome resulted in per-sample average alignment rates ranging from 80.6 – 81.6% (Additional File S2). Genotyping used 408.9 M alignments, of which 367.9 M (90%) were primary alignments and were retained. Low quality (6%) or excessively soft clipped alignments (4%) were removed. Genotyping identified 383,371 unfiltered loci with an average of 80.7 sites per locus and an average per-sample effective coverage of 91.5 ± 28.7x (range: 2.9-176.6x). Filtering to remove loci with excess missing data (see *Methods*) or low global minor allele frequency (i.e., MAF < 0.01) resulted in the removal of 370,235 loci (96.6%). Within the retained 13,136 filtered loci, comprising 1.14 M genomic sites, 21,048 variants were identified within 9,377 polymorphic RAD-tags.

When retaining multiple variants per RAD-tag, there was an average of 2.24 variants within each tag, and a single, two, or three variants were observed in 3,721 (39.7%), 2,585 (27.6%), and 1,541 (16.4%) tags, respectively (Additional File S3). Four to nine variants were observed in 1,518 tags (16.1%), and 10-13 variants per tag were observed in 12 tags. Counts of two, three, or four alleles per tag were observed in 74% of the tags (i.e., 4,073, 2,609, 1,431 RAD-tags, respectively; Additional File S3). Five to nine alleles per tag were observed in 1,215 tags (23.8%), 45 tags had between 10-16 alleles, and four had over 20 alleles each. Notably, these tallies include variants not yet filtered for deviations from Hardy-Weinberg (HW) proportions.

Per individual missing data was on average 8.1 ± 12.2%. Samples with more than 30% missing data were dropped from the analysis, which removed one sample from VIU and 13 samples from JPN (see Table 1 for sample size post filters). After filtering on missing data, 173 individuals remained, with an average per sample missing data of 5.0 ± 4.3%. Of the 9,375 variants in the single variant per RAD-tag

dataset, 2,578 (27.5%) did not conform to expected HW proportions ($p < 0.01$) in at least one population and were removed. For the BC farm, JPN, and VIU collections this included 703, 1,773, and 1,015 non-conforming markers, respectively. Of the remaining 6,797 variants, 74 had excess heterozygosity (i.e., $H_{OBS} > 0.5$) and were removed, leaving 6,723 variants remaining. As samples had been removed, a second MAF filter was applied, and this resulted in the removal of an additional 176 variants with MAF < 0.01. After all filters, the single variant per RAD-tag dataset had 6,547 variants remaining for downstream analyses.

3.2. Global genetic differentiation

In an unsupervised PCA using the single-SNP per locus data (n = 6,547 variants), samples separated by country (i.e., Canada or Japan) across PC1, explaining 6.5% of the overall variation (Figure 1A). The BC farm and VIU samples separated across PC2 (5.2% of the variance explained). PC3 and PC4 each explained 3.3% and 3.0% of the variation, respectively, and these axes separated samples within each of the BC farm and VIU populations (Figure 1B). After PC4, individual PCs were less informative (Figure 1A inset). Genetic differentiation analysis using F_{ST} (Weir & Cockerham 1984) indicates significant global genetic differences between all of the populations (Table 2). Similar genetic differentiation was observed between the VIU hatchery and the JPN wild population (95% CI F_{ST} = 0.067-0.074) as between the two cultured populations (F_{ST} = 0.076-0.085). The highest differentiation was between the BC farm and the JPN wild population, with F_{ST} = 0.093-0.101.

3.3. Genetic diversity, low frequency variants, and private alleles

Considering all 21,048 variants (i.e., multiple SNPs per RAD-tag) prior to HW filters, the JPN collection had the highest polymorphism rate at 1.7%, with 19,411 variants (Table 3). By comparison, there were 15,471 variants in the VIU collection (i.e., 1.35% polymorphism rate), and 12,286 variants in the BC farm collection (i.e., 1.07% polymorphism rate). However, observed heterozygosity (H_{OBS}) was higher in the BC farm and VIU populations than JPN with mean H_{OBS} estimated at 0.00278, 0.00276, and 0.00266, respectively (S.E. for all: ± 0.00003). Using the rarefied dataset (n = 37 samples per population), these relative trends in polymorphism rate and H_{OBS} remained consistent (Table 3). Therefore, the highest polymorphism rate but lowest mean H_{OBS} is observed in the wild samples (JPN) relative to the cultured samples, and the BC farm collection has the lowest polymorphism rate.

Using the filtered, single-SNP per RAD-tag dataset, 4,401 of the total 6,547 variants (67.2%) had MAF ranging 0.01-0.10. Once the data were separated by individual population, and any monomorphic or low MAF variants were removed (i.e., population MAF < 0.01), the JPN, VIU, and BC farm datasets retained 5,423 (82.8%), 4,531 (69.2%), and 3,722 (56.9%) variants, respectively (Figure 2). In addition to having the most variants retained, JPN also had the highest proportion of retained lower frequency variants

289 (i.e., MAF between 0.01-0.10) with 61.8% of the variants in this range, followed by VIU (49.7%), and
290 lastly the BC farm (39.8%).
291 Private alleles within the single-SNP per locus filtered dataset were also most numerous in JPN
292 with 1,401 private alleles, followed by VIU (n = 270), and BC farm (n = 58). Regionally, when VIU and
293 BC farm were merged into a Canadian population then compared to JPN, there were 1,401 private alleles
294 in JPN and 690 in Canada. Although less numerous than JPN, the Canadian private alleles were higher
295 overall in MAF (median MAF = 0.032) than the private alleles in JPN (median MAF = 0.018; Figure 3).
296 Effective population size (N_e) was estimated for each population using either the microhaplotype
297 multiple-SNP per RAD-tag dataset (n = 21,048 variants, 9,377 RAD-tags) or the single-SNP per RAD-tag
298 dataset (n = 9,375 variants). The linkage disequilibrium (LD) method estimated 95% CI of N_e for JPN as
299 25,048 – 56,291 (n = 9,004 polymorphic tags), for VIU as 26.4 – 26.4 (n = 7,926 tags), and for BC farm as
300 9.3 – 9.3 (n = 6,839 tags) when using a P_{CRIT} value of 0.01 (Table 3). The heterozygosity excess method
301 was unable to estimate the 95% CI for variants with MAF > 0.01, and the molecular coancestry method
302 estimates were very low for all populations (estimated N_{eb} 2.6 – 6.5; see Additional File S4). Using the
303 single-SNP per RAD-tag data, the 95% CI estimates of N_e with the LD method was not calculable using
304 the data for JPN (lower and upper values were both infinite) but was estimated at similar levels for VIU
305 and BC farm as the microhaplotype data (i.e., 26.2 – 26.2 for VIU and 9.2 – 9.3 for BC farm). Using the
306 rarefied dataset (n = 37 individuals per population), 95% CI of N_e for VIU and BC farm were again
307 estimated to be 9.5 and 22.9-23.0, respectively, but the JPN population N_e was not estimable with the
308 existing data (Additional File S4).
309

310 **3.4. Inter-individual relatedness**

311 Using the microhaplotype data in fineRADstructure (Malinsky *et al.* 2018), clusters of related individuals
312 were observed in the VIU and BC farm collections, whereas the JPN collection was comprised of
313 approximately equally dissimilar samples (Additional File S5). The BC farm has three main clusters
314 containing 20, 12, and four individuals each; these were the same groupings as clusters observed across
315 PC3 and PC4 (Figure 1B; Additional File S6). The VIU collection has many small clusters of related
316 individuals, and some of these individual clusters are estimated to be more different from each other than
317 are the clusters in the BC farm, based on shared microhaplotypes. Within the JPN collection, slight
318 clustering is observed in a grouping of seven individuals, and a grouping of 11 individuals. The cause of
319 these slight groupings is not clear given that it encompasses both marine survey and commercial harvest
320 samples. When considering relatedness using the single-SNP per locus dataset, the JPN collection is
321 observed to have a relatively tight distribution of relatedness compared to the VIU or BC farm collections
322 (Figure 4). Many highly related outliers are observed in the VIU collection, as expected given the groups

323 of related individuals in the microhaplotype analysis for this collection. One JPN sample (JPN_110) showed
324 high relatedness to BC_029, and the cause of this is unknown, given that each sample also show high
325 similarity to other members of their own collections.
326

327 **3.5. Per locus heterozygosity and highly variable tags**

328 Filtered single-SNP per RAD-tag markers were inspected for per-locus H_{OBS} within individual populations
329 of JPN or VIU, then values were compared to identify variants that are expected to have high heterozygosity
330 in both collections (Figure 5; Additional File S7 for per-locus values). Additionally, given the observation
331 of RAD-tags with high numbers of variants per tag, the marker names, the number of variants per tag, and
332 whether the tag is within HW proportions is shown in Additional File S7. The single-SNP per locus and
333 microhaplotype dataset VCF files are provided, which gives the location of the markers in the reference
334 genome (see *Data Availability*).
335

336 # Discussion

337 The present study was conducted to evaluate the degree to which cultured Yesso scallop populations in
338 British Columbia have lost genetic diversity since their founding in the 1980s. This benchmark analysis
339 provides a starting point for continued monitoring and preservation of existing genetic diversity on an
340 ongoing basis. Although the standard hatchery practice of crossing divergent lines to avoid inbreeding and
341 to gain heterosis benefits is observed in the present data as elevated heterozygosity in cultured stocks, low
342 effective population sizes and significant genetic diversity decreases indicate that it is important to continue
343 efforts to maintain, monitor, and optimally use the existing diversity present in the cultured populations.
344

345 *Genomic diversity in wild Yesso scallop from northern Japan*

346 Wild scallops from northern Japan (n = 71) were estimated to have a 1.7% polymorphism rate when
347 including all SNPs with MAF > 0.01 that were present in at least 70% of the samples from each of three
348 populations and was calculated prior to removing variants that did not conform to Hardy-Weinberg
349 proportions. This rate was estimated based on 0.08% (or 1.14 Mbp) of the expected 1.43 Gbp genome
350 (Wang *et al.* 2017). Although shellfish polymorphism rates are expected to be high (Plough 2016), this rate
351 is higher than that calculated for the wild, hermaphrodite parent individual used for the reference genome
352 assembly, which was estimated at 1.04% (Wang *et al.* 2017). The estimated rate in the present study was
353 also higher than that observed in the Pacific oyster genome, which was evaluated at 1.30% (Zhang *et al.*
354 2012). The king scallop *Pecten maximus* was estimated to have a 1.7% heterozygosity rate for the sequenced
355 individual, which was higher than that estimated in the Pacific oyster assembly, for which the authors

356 ascribe to the breeding program source of the sequenced oyster (Kenny *et al.* 2020). Sutherland and co-
357 workers (2020) in an analysis of a variety of wild, naturalized, and farmed Pacific oysters identified a mean
358 polymorphism rate of 1.43%, and lower rates in farmed populations (e.g., 0.94%-1.13%). The high
359 polymorphism rate in Pacific oyster, also estimated at 1 SNP per 40 bp (2.5%), was given as one of the
360 reasons for pursuing the reference genome, as well as a challenge to overcome in order to do so (Hedgecock
361 *et al.* 2005). Combining the high polymorphism rate with sweepstakes reproductive success (SRS) effects
362 on annual variation in allele frequencies leads to the question of what the longevity is of many of the variants
363 identified here in the lower frequency range (i.e., 62% of variants with MAF between 0.01 and 0.10).
364 The high polymorphism rate resulted in some RAD-tags having a numerous variants within the
365 same 80 bp fragment, where 16% of tags had more than four variants. These highly variable regions may
366 need to be treated with caution, for example if used for marker design, since the genome assembly may
367 contain collapsed segments, given the differences between the expected and the assembled genome size
368 (Wang *et al.* 2017). High repeat content resulting in missing genomic segments was also an issue for the
369 Pacific oyster (Zhang *et al.* 2012), which is thought to be comprised of 35% repetitive elements (Hedgecock
370 *et al.* 2005).
371 The effective population size (N_e) of the wild Japan collection characterized here was estimated to
372 be between 25-56K breeding individuals. The census size of the population around Hokkaido is not known,
373 but approximately 145-375K individuals are expected to have been harvested annually in the bay from
374 which samples were collected (Uchiura Bay, N. Itoh *pers. comm.*).
375
376 *Cultured strains and genomic impacts of breeding programs*
377 The BC cultured scallop populations characterized here have significantly lower polymorphism rates than
378 the source population. The genotyped VIU collection, which likely represents the full complement of
379 genetic backgrounds in the VIU breeding program (T. Green, *pers. obs.*), was estimated to have a
380 polymorphism rate of 1.35%, which is 21% lower than the source wild population. The BC farm is likely
381 only a subset of the full genetic variation in the cultured lineages from the seed provider for the farm since
382 this collection was obtained as farmed adults from a single sampling event. This collection had a 1.07%
383 polymorphism rate, which is 59% lower than the source population. Both cultured collections were also
384 typified by a depletion of alleles in the lower MAF range, where a MAF range of 0.01 to 0.10 comprised
385 62% of the variants in JPN, but only 50% and 40% in VIU and BC farm, respectively).
386 Genetic changes from source populations are also observed as skewed allele frequencies genome-
387 wide, which can occur through founder effects and drift. For example, Suminoe oyster hatchery strains
388 were differentiated both against the source population and between each strain (F_{ST} = 0.05-0.24) (Xiao *et*
389 *al.* 2011). These impacts are likely due to few parents being included in the original founding population

390 (i.e., founder effects) and drift occurring due to a low effective population size, but potentially also from
391 any artificial selection that has occurred (Li *et al.* 2007; Xiao *et al.* 2011). Hatchery strains of the common
392 blue mussel *Mytilus edulis* were found to be significantly different from each other (F_{ST}: 0.03-0.08) as well
393 as from the source wild population (F_{ST} = 0.07-0.08), which accompanies reduced N_e and allelic diversity
394 (Gurney-Smith *et al.* 2017). Eastern oyster hatchery strains also show significant differentiation (F_{ST} =
395 0.076) between hatchery and wild populations (Carlsson *et al.* 2006), and eastern oyster strains increase in
396 their differentiation within and between hatchery strains over time (Varney & Wilbur 2020). Pacific oyster
397 hatchery strains showed significant differentiation from source populations, dwarfing the differentiation
398 observed between natural populations in different countries along the same translocation lineage
399 (Sutherland *et al.* 2020). In Yesso scallop, here we observed high differentiation between hatchery lineages
400 as well as between both hatchery lineages with the source population (F_{ST}: 0.07-0.10). As a notable contrast,
401 five different farms characterized in China showed generally low differentiation between farms (F_{ST} = 0.02-
402 0.03), but high differentiation between the farms and the source populations (F_{ST} = 0.09-0.15) (Li *et al.*
403 2007). It is possible that this similarity between the farms in China may indicate a common seed scallop
404 provider for the different farms. In general, the cultured lineages here show the expected level of
405 differentiation from the wild population relative to these other shellfish species, many of which were
406 referenced as a concern in the original studies and needing to be monitored for continued diversity losses.

407 The N_e of the cultured lineages here were significantly reduced relative to the wild population,
408 indicating founder effects and drift. The N_e estimated for the VIU population was 26, conforming to the
409 expectations based on the known number and diversity of the source parents used to initiate this breeding
410 population (T. Green, *pers. obs.*). The BC farm population, which may not fully represent the entire
411 broodstock, as described above, was estimated at N_e = 9. These values are remarkably lower than that
412 observed for the wild population (i.e., N_e = 25-56K), which required all data to estimate, and was only
413 estimable using microhaplotype information. Similarly, cultured stocks of Yesso scallop analyzed by Li
414 and co-workers (2007) had N_e calculated between 27-70, even though hundreds of males and females are
415 used to propagate the hatchery strains. These levels were a concern regarding the long term sustainability
416 of the stock (Li *et al.* 2007). Similarly, Gurney-Smith and co-workers (2017) estimated N_e values under 50
417 for the three hatchery strains of common blue mussel analyzed, and the wild population was too large to
418 estimate with the available data. Considering the genome-wide scope of the allele frequency changes, and
419 the significant decreases in N_e, it is likely that founder effects and drift are sufficient to explain the allele
420 frequency shifts observed in the present study.

421 An additional indicator of founder effects and drift is that private alleles in the cultured lineages
422 were of relatively high frequency (MAF = 0.01-0.28) relative to the wild population. Although these private
423 alleles are likely to be in the Japan wild population, they were not represented in the sampling here, and

424 therefore not expected at the same high frequency as those in culture. By contrast, wild private alleles,
425 although more numerous, were lower in frequency (MAF: 0.01-0.15).

426 Seemingly paradoxical to known inbreeding expectations, with an initial decrease in polymorphism
427 and a subsequent loss of heterozygosity expected (Hedgecock & Sly 1990), here the cultured Yesso scallops
428 had elevated H_{OBS} relative to the wild population. This is likely due to the shellfish breeding practice of
429 outcrossing, or crossing inbred lines, to produce heterosis in the offspring (Hedgecock *et al.* 1995;
430 Hedgecock & Davis 2007). Heterosis is particularly valuable in shellfish due to the high genetic load in the
431 taxon, as discussed above. Notably, high observed heterozygosity and therefore low inbreeding coefficient
432 was observed alongside the lowest polymorphism rate in a farmed population of Pacific oyster, likely to be
433 explained by the same breeding approaches for heterosis as observed here (Sutherland *et al.* 2020). Selected
434 strains of eastern oyster also showed elevated H_{OBS} in the hatchery strains (Carlsson *et al.* 2006; Varney &
435 Wilbur 2020), which the authors attribute to artificial selection and increased frequency of the remaining
436 alleles replacing low frequency alleles lost due to drift and founder effects (Hedgecock & Sly 1990; Varney
437 & Wilbur 2020). Whether H_{OBS} will be elevated or reduced in hatchery strains relative to the wild
438 populations will depend on hatchery practices; reduced H_{OBS} relative to wild populations (11% lower) was
439 observed in the longer established Suminoe oyster (Xiao *et al.* 2011). Given these varying trends in H_{OBS}
440 in the above cases, all of which have a consistent decreasing trend in genetic diversity by other metrics, and
441 given that decreases in H_{OBS} can be counteracted relatively quickly through crossing of divergent lines
442 (Hedrick 2005), as well as the fact that allelic diversity will be the first to decrease (Xiao *et al.* 2011), loss
443 of alleles is likely a better metric for monitoring diversity loss in cultured bivalve populations than
444 heterozygosity or inbreeding coefficients (Yu & Guo 2004; Varney & Wilbur 2020).

445 Breeding programs in general are challenged by losses of genetic variation over successive
446 generations of breeding with or without selection, which in shellfish is further compounded by high
447 fecundity and large differences in reproductive success, resulting in significant drift (Boudry *et al.* 2002;
448 Varney & Wilbur 2020). This problem becomes even more challenging if the source population is not
449 available for replenishing the broodstock through introgression, as is the case when the cultured stock is
450 grown outside of its native range (Li *et al.* 2007; Gurney-Smith *et al.* 2017). This is unfortunately the
451 situation for Yesso scallop in British Columbia, where no naturalized populations exist outside of hatchery
452 breeding programs. This increases the need for effective management of shellfish hatcheries including
453 genetic diversity monitoring (Gurney-Smith *et al.* 2017), good record keeping, and the availability of
454 genetic tools to confirm pedigrees, low relatedness of parents in target crosses, and confirmed identity of
455 lineage for the individuals chosen as broodstock for crosses (Hedgecock & Davis 2007).

456

Conclusions

The wild Yesso scallop characterized from Japan has a high polymorphism rate similar to other shellfish species, and therefore likely will be significantly impacted by both inbreeding depression and heterosis. Cultured populations characterized from the industry in BC indicate strong founder effects and drift, with effective population sizes around 9-26 individuals relative to that calculated from the wild population at 25-56K individuals, a general depletion of low frequency variants in cultured populations and high frequency private alleles, and high F_{ST} values between cultured and the wild population. The polymorphism rate was reduced in the cultured populations, but observed heterozygosity was elevated, likely due to the practice of outcrossing to induce heterosis, indicating that polymorphism rate is a better estimator for genetic diversity loss than observed heterozygosity in this species. Although significant diversity loss is observed relative to wild populations, existing cultured populations still have relatively high polymorphism rates, and efforts to monitor preserve the standing genetic variation will be important to continue to ensure the long-term viability of the broodstock program and the adaptability of the cultured animals to environmental challenges and selective breeding efforts.

Data Availability

Code and README required for the analysis: https://github.com/bensutherland/ms_scallop_popgen

Analysis pipeline applied: https://github.com/enormandeau/stacks_workflow

Raw sequencing data has been uploaded to SRA under BioProject PRJNA947158, BioSamples SAMN33843243-SAMN33843431.

Additional material is included on figshare, including the Stacks sample information file required for *stacks_workflow*, and various file formats for single-SNP per locus or microhaplotype data required for the analysis: doi.org/10.6084/m9.figshare.22670626

481 **Additional Files**

482 Additional File S1. Sample metadata including sources and dates of collections.

483 Additional File S2. Per sample number of reads, alignments, and alignment rate.

484 Additional File S3. Number of variants and alleles per RAD-tag, including summary tables.

485 Additional File S4. Effective population size (N_e) results using microhaplotypes or single variant per tag.

486 Additional File S5. Inter-individual relatedness as calculated by microhaplotypes in fineRADstructure.

487 Additional File S6. Per sample principal component (PC) scores for PC1-4.

488 Additional File S7. Per locus stats including F_{ST}, H_{OBS} (global) for each locus in the filtered, single-SNP

489 per locus dataset, H_{OBS} in each population, and Hardy-Weinberg equilibrium test outputs for single SNPs

490 or microhaplotypes per population.

References

Abed A, Légaré G, Pomerleau S, St-Cyr J, Boyle B, Belzile FJ (2019) Genotyping-by-sequencing on the Ion Torrent platform in barley. In: *Barley: Methods and Protocols* (ed. Harwood WA), pp. 233-252. Springer New York, New York, NY.

Andrews S (2010) FastQC: A quality control tool for high throughput sequence data.

Anisimova A (2007) Genome sizes of some Bivalvia species of the Peter the Great Bay of the Sea of Japan. *Comparative Cytogenetics* **1**, 63-69.

Boudry P, Collet B, Cornette F, Hervouet V, Bonhomme F (2002) High variance in reproductive success of the Pacific oyster (*Crassostrea gigas*, Thunberg) revealed by microsatellite-based parentage analysis of multifactorial crosses. *Aquaculture* **204**, 283-296.

Bourne NF (2000) The potential for scallop culture–the next millenium. *Aquaculture International* **8**, 113-122.

Bower SM, Blackbourn J, Meyer GR, Welch DW (1999) Effect of *Perkinsus qugwadi* on various species and strains of scallops. *Diseases of Aquatic Organisms* **36**, 143-151.

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

Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, Handsaker RE, Lunter G, Marth GT, Sherry ST, McVean G, Durbin R, Group GPA (2011) The variant call format and VCFtools. *Bioinformatics* **27**, 2156-2158.

Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, Pollard MO, Whitwham A, Keane T, McCarthy SA, Davies RM, Li H (2021) Twelve years of SAMtools and BCFtools. *GigaScience* **10**.

DFO (2023) Aquaculture production quantities and value, https://www.dfo-mpo.gc.ca/stats/aqua/aqua-prod-eng.htm.

Do C, Waples RS, Peel D, Macbeth GM, Tillett BJ, Ovenden JR (2014) NeEstimator v2: re-implementation of software for the estimation of contemporary effective population size (Ne) from genetic data. *Molecular Ecology Resources* **14**, 209-214.

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.

Ewels P, Magnusson M, Lundin S, Käller M (2016) MultiQC: summarize analysis results for multiple tools and samples in a single report. *Bioinformatics* **32**, 3047-3048.

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

Gregory TR (2022) Animal Genome Size Database.

Gurney-Smith HJ, Wade AJ, Abbott CL (2017) Species composition and genetic diversity of farmed mussels in British Columbia, Canada. *Aquaculture* **466**, 33-40.

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.

Hedrick PW (2005) Inbreeding. In: *Genetics of populations*, pp. 237-292. Jones and Bartlett Publishers, Sudbury, Massachusetts.

Holden JJ, Collicutt B, Covernton G, Cox KD, Lancaster D, Dudas SE, Ban NC, Jacob AL (2019) Synergies on the coast: challenges facing shellfish aquaculture development on the central and north coast of British Columbia. *Marine Policy* **101**, 108-117.

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.

Kenny NJ, McCarthy SA, Dudchenko O, James K, Betteridge E, Corton C, Dolucan J, Mead D, Oliver K, Omer AD, Pelan S, Ryan Y, Sims Y, Skelton J, Smith M, Torrance J, Weisz D, Wipat A, Aiden EL, Howe K, Williams ST (2020) The gene-rich genome of the scallop *Pecten maximus*. *GigaScience* **9**.

Kobayashi T, Kijima A (2010) Effects of inbreeding depression in Pacific abalone *Haliotis discus hannai*. *Journal of Shellfish Research* **29**, 643-649.

Lepais O, Weir JT (2014) SimRAD: an R package for simulation-based prediction of the number of loci expected in RADseq and similar genotyping by sequencing approaches. *Molecular Ecology Resources* **14**, 1314-1321.

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

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.

Malinsky M, Trucchi E, Lawson DJ, Falush D (2018) RADpainter and fineRADstructure: Population Inference from RADseq Data. *Molecular Biology and Evolution* **35**, 1284-1290.

Martin M (2011) Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal* **17**, 3.

Mascher M, Wu S, Amand PS, Stein N, Poland J (2013) Application of genotyping-by-sequencing on semiconductor sequencing platforms: a comparison of genetic and reference-based marker ordering in barley. *PLOS ONE* **8**, e76925.

Meyer GR, Lowe GJ, Gilmore SR, Bower SM (2017) Disease and mortality among Yesso scallops *Patinopecten yessoensis* putatively caused by infection with *Francisella halioticida*. *Diseases of Aquatic Organisms* **125**, 79-84.

Paradis E (2010) pegas: an R package for population genetics with an integrated–modular approach. *Bioinformatics* **26**, 419-420.

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.

Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, Maller J, Sklar P, de Bakker PI, Daly MJ, Sham PC (2007) PLINK: a tool set for whole-genome association and population-based linkage analyses. *American Journal of Human Genetics* **81**, 559-575.

Purcell SL (2022) PLINK (v.1.90b6.26).

R Core Team (2023) R: A Language and Environment for Statistical Computing, Vienna, Austria.

Rochette NC, Rivera-Colón AG, Catchen JM (2019) Stacks 2: Analytical methods for paired-end sequencing improve RADseq-based population genomics. *Molecular Ecology* **28**, 4737-4754.

Solé-Cava AM, Thorpe JP (1991) High levels of genetic variation in natural populations of marine lower invertebrates. *Biological Journal of the Linnean Society* **44**, 65-80.

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.

Tange O (2022) GNU Parallel 20220822. Zenodo.

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.

Wang S, Zhang J, Jiao W, Li J, Xun X, Sun Y, Guo X, Huan P, Dong B, Zhang L, Hu X, Sun X, Wang J, Zhao C, Wang Y, Wang D, Huang X, Wang R, Lv J, Li Y, Zhang Z, Liu B, Lu W, Hui Y, Liang J, Zhou Z, Hou R, Li X, Liu Y, Li H, Ning X, Lin Y, Zhao L, Xing Q, Dou J, Li Y, Mao J, Guo H, Dou H, Li T, Mu C, Jiang W, Fu Q, Fu X, Miao Y, Liu J, Yu Q, Li R, Liao H, Li X, Kong Y, Jiang Z, Chourrout D, Li R, Bao Z (2017) Scallop genome provides insights into evolution of bilaterian karyotype and development. *Nature Ecology & Evolution* **1**, 0120.

Weir BS, Cockerham CC (1984) Estimating F-Statistics for the Analysis of Population Structure. *Evolution* **38**, 1358-1370.

Williams GC (1975) *Sex and Evolution. (MPB-8), Volume 8* Princeton University Press.

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.

Zhao L, Li Y, Li Y, Yu J, Liao H, Wang S, Lv J, Liang J, Huang X, Bao Z (2017) A genome-wide association study identifies the genomic region associated with shell color in Yesso scallop, *Patinopecten yessoensis*. *Marine Biotechnology* **19**, 301-309.

Zhao L, Li Y, Lou J, Yang Z, Liao H, Fu Q, Guo Z, Lian S, Hu X, Bao Z (2019) Transcriptomic profiling provides insights into inbreeding depression in Yesso scallop *Patinopecten yessoensis*. *Marine Biotechnology* **21**, 623-633.

Figures

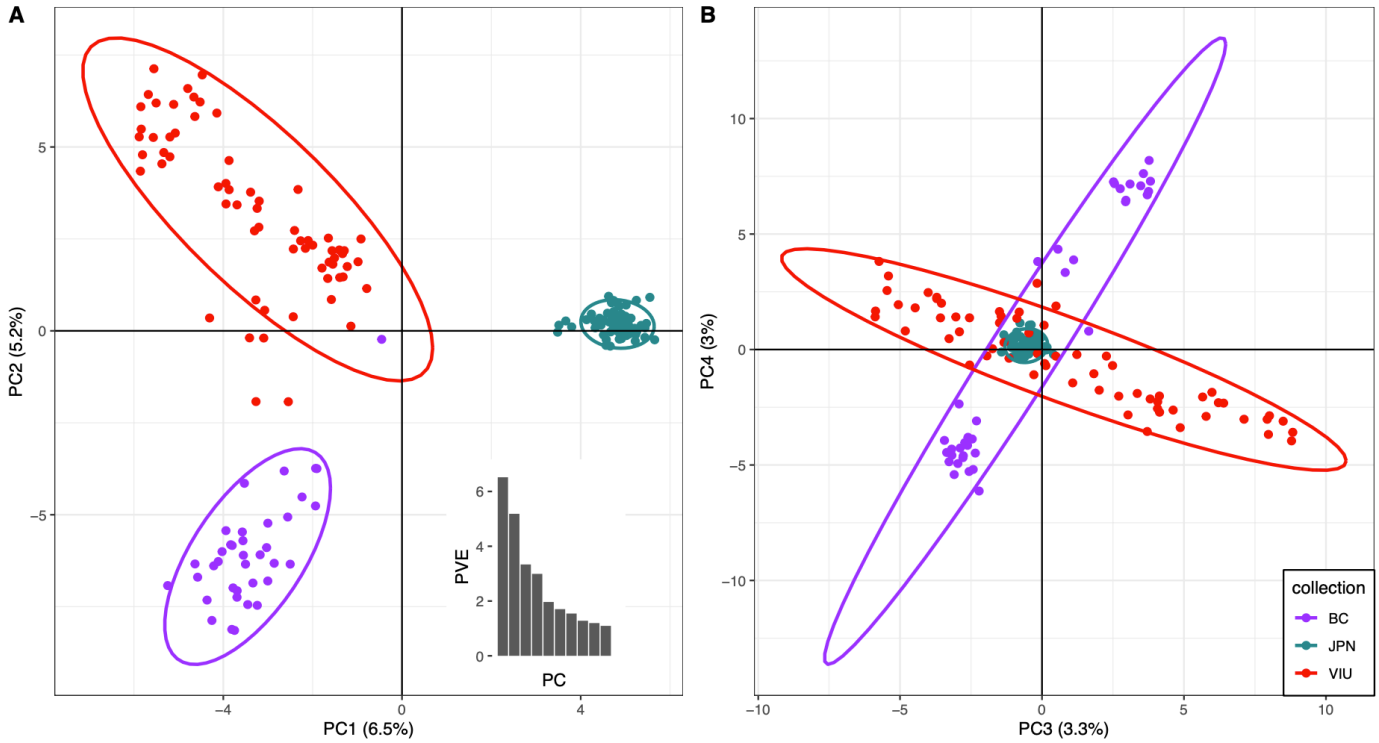

Figure 1. Principal components analysis (PCA) separates the three collections. (A) Farmed or breeding program scallops in British Columbia (BC), Canada are separated from wild scallops in Japan across PC1 (6.5% variance explained). PC2 separates the two BC collections (5.2% variance explained). (B) PC3 and PC4 separate variation within BC collections. After the first four PCs, the percent variation explained reduces as shown in the scree plot (see inset, panel A).

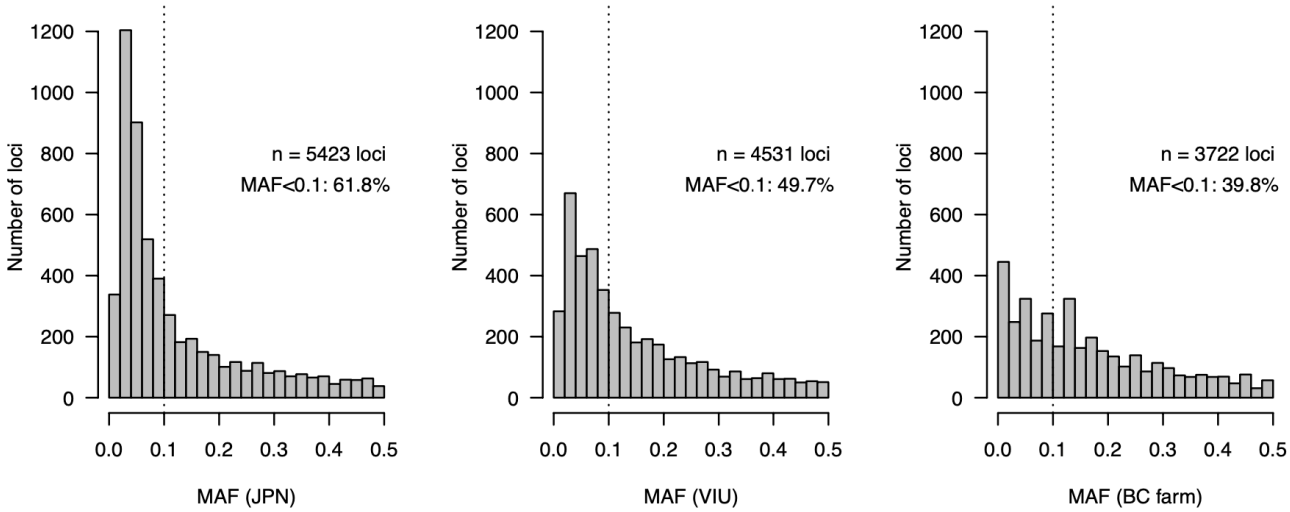

Figure 2. Per locus minor allele frequency (MAF) distribution for the Japan wild population, the VIU breeding population, and the BC farm population following all filters, including population-specific low MAF filters.

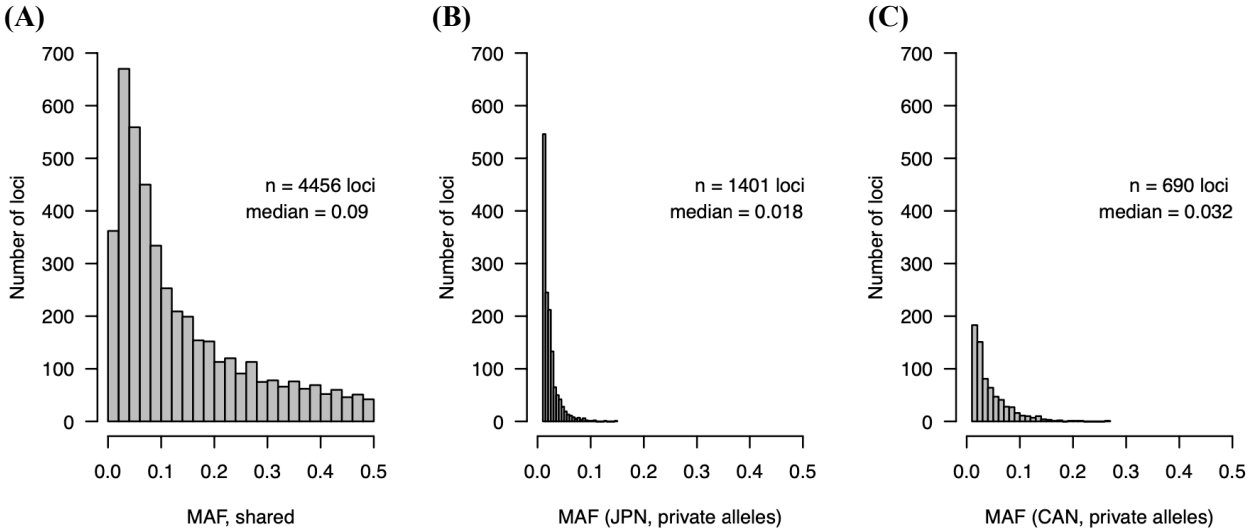

Figure 3. MAF distributions for (A) loci found in all populations; (B) private alleles in Japan; and (c) private alleles in Canada (VIU and BC farm).

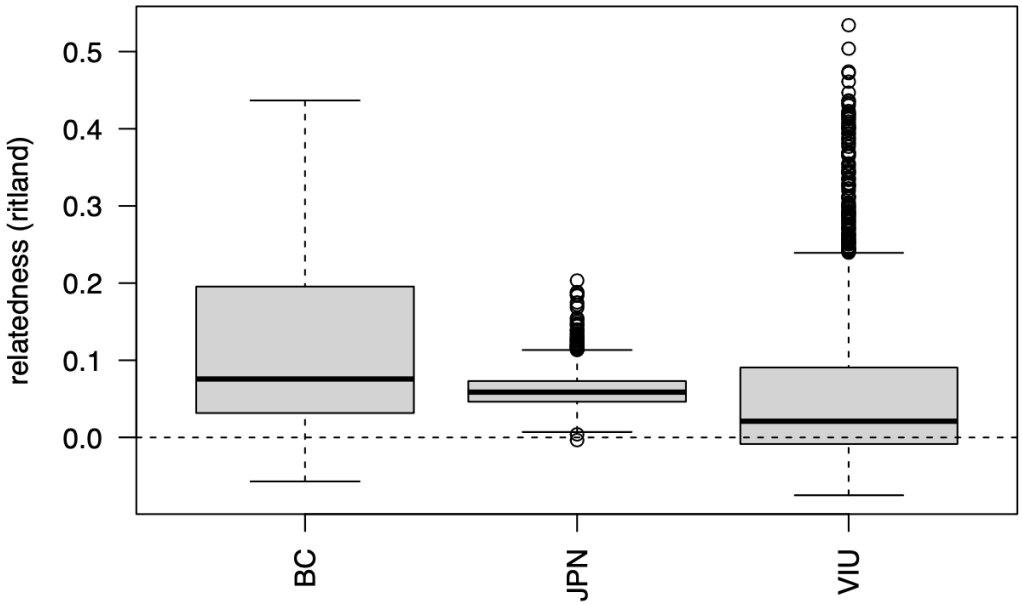

Figure 4. Relatedness within each population as evaluated by the Ritland statistic, using a single SNP per locus.

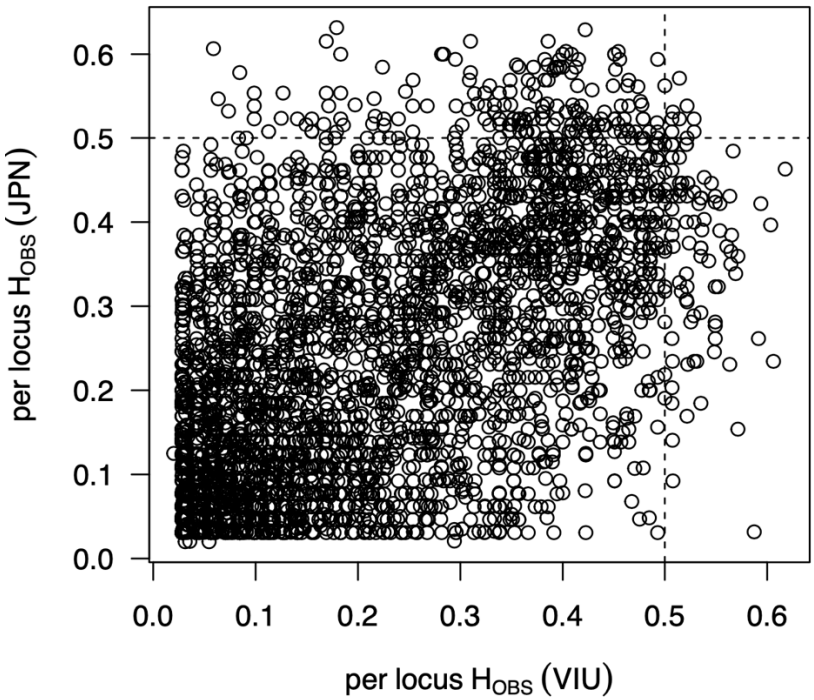

Figure 5. JPN and VIU population-specific H_{OBS} comparison using the filtered, single-SNP per locus dataset.

Tables

Table 1. Overview of genotyped samples including sources, sample size per population (n), total filtered and demultiplexed reads (M), average ± s.d. reads and alignments per population, and samples passing all filters.

Collection	Source	n	Total reads (M)	Reads/ sample (M)	Alignments/ sample (M)	Alignment rate (%)	n (post filters)
BC	BC farm source near Vancouver Island	37	127.5	3.45 ± 0.93	2.79 ± 0.77	80.9 ± 1.5	37
JPN (1)	Wild, survey	59	100.5	1.7 ± 0.60	1.40 ± 0.51	80.5 ± 2.6	45
JPN (2)	Wild, commercial, North Hokkaido	26	91.0	3.5 ± 1.20	2.82 ± 0.98	80.5 ± 1.4	26
VIU	Cultured, breeding program	67	184.5	2.80 ± 1.07	2.29 ± 0.92	80.5 ± 9.3	65
Total/ All	**-**	**189**	**503.5**	**2.69 ± 1.17**	**2.19 ± 0.97**	**80.6 ± 5.8**	**173**

Table 2. Population genetic differentiation (F_{ST}) comparison expressed as 95% confidence intervals with the upper limit on the upper right and lower limit in the lower left. Calculations used the single-SNP per locus data after filters.

	BC	**JPN**	**VIU**
BC	-	0.101	0.085
JPN	0.093	-	0.074
VIU	0.076	0.067	-

Table 3. Summary statistics from genotyping with the full dataset or the rarefied dataset. Both datasets allow multiple variants per RAD-tag, and the full dataset has 21,048 variants in 1,143,426 sites, the rarefied had 20,269 variants in 1,746,980 genomic sites. Number of polymorphic sites and polymorphism rate (%) are shown. Observed heterozygosity (H_{OBS}) is calculated based on all positions (variant and fixed). Effective population size (N_e) was calculated based on microhaplotypes (9,377 loci). Filtered loci shows the number of polymorphic loci per population in single-SNP per locus data (n = 9,375 variants before filters, n = 6,547 variants) retained after all filters, including Hardy-Weinberg equilibrium.

Dataset	**Pop**	**Samples / locus**	**Poly. sites**	**Poly. rate (%)**	**H_{OBS}**	**N_e (95% CI)**	**Filtered loci**	**Private alleles**
Full	BC	35.9	12,286	1.07	0.00278	9.3 – 9.3	3,722	58
	JPN	74.7	19,411	1.70	0.00266	25,048 – 56,291	5,857	1,401
	VIU	63.7	15,471	1.35	0.00276	26.4 – 26.4	4,856	270
Rarefied	BC	34.2	12,685	0.73	0.00186	9.5 – 9.5	-	-
	JPN	35.7	17,490	1.00	0.00177	∞	-	-
	VIU	35.4	14,765	0.85	0.00184	22.9 – 23.0	-	-