

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

Ben J.G. Sutherland <sup>1,2,\*</sup> Naaki Itoh,<sup>3</sup> Korrina Gilchrist,<sup>1</sup> Brian Boyle,<sup>4</sup> Myron Roth,<sup>5</sup> Timothy J. Green<sup>1</sup>

<sup>1</sup>Faculty of Science and Technology, Vancouver Island University, Nanaimo, British Columbia V9R 5S5, Canada

<sup>2</sup>Sutherland Bioinformatics, Lantzville, British Columbia V0R 2H0, Canada

<sup>3</sup>Laboratory of Fish Diseases, Graduate School of Agricultural and Life Sciences, The University of Tokyo, Tokyo 113-8657, Japan

<sup>4</sup>Plateforme d'Analyses Génomiques, Institut de Biologie Intégrative et des Systèmes (IBIS), Université Laval, Québec, Québec G1V 0A6, Canada

<sup>5</sup>BC Ministry of Agriculture and Food, Victoria, British Columbia V8W 9B4, Canada

\*Corresponding author: Faculty of Science and Technology, Vancouver Island University, 900 Fifth St, Nanaimo, BC V9R 5S5. Email: Sutherland.Bioinformatics@protonmail.com

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 ranged 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 was 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 to 26 individuals whereas the wild population was estimated at 25,048–56,291 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 ( $H_{OBS}$ ; Solé-Cava and 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). The high polymorphism rate may also be related to transposable element activity (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 ( $N_e$ ) relative to census size from sweepstakes reproductive success (SRS; Hedgecock 1994; Hedgecock and 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 and Davis 2007).

The Yesso scallop *Mizuhopecten yessoensis* is an economically important aquaculture species and a valuable commercial

protein source. Native to Japan (JPN), 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 tons per year, but BC has not reached its production potential, despite favorable environmental conditions (Holden et al. 2019). BC production of Yesso scallop averaged 89 tons per annum in 2017–2020 but decreased to 52 tons 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 2–3 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 halitoides* or the protozoan *Perkinsus qugwadi* (Bower et al. 1999; Bourne 2000; Meyer et al. 2017). The industry

is therefore now looking toward 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 and Sly 1990; Evans et al. 2004; Gurney-Smith et al. 2017). For example, hatchery lineages of Asian Suminoe oyster *Crassostrea 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 and Wilbur 2020). In the aforementioned cultured populations of Yesso scallop in China, reduced polymorphism rate and heterozygosity have occurred, which are major concerns 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 and 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 and Komaru 1994), Pacific oyster *Crassostrea gigas* (Evans et al. 2004), and Pacific abalone *Haliotis discus hannai* (Kobayashi and 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 coloration (Zhao et al. 2017), efforts must also be put toward 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 toward 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 JPN through commercial harvest or research surveys. Comparisons between these populations will indicate the current state of the assessed cultured lineages in BC.

## Materials and methods

### Sample collections

*Mizuhopecten yessoensis* samples were obtained from 4 different sources (Table 1). Wild scallops from JPN were obtained from marine surveys in Southern Hokkaido in late 2017 through early 2021 (see Supplementary File S1) and from a commercial source as harvested wild scallops from Northern Hokkaido in 2021. BC-farmed scallops were also obtained from a commercial provider that were grown near central Vancouver Island, BC. Samples from the VIU Centre for Shellfish Research breeding program were obtained from animals in late 2022.

### DNA extraction, quantification, and quality control

Marine survey samples were obtained as pelleted DNA in 100% ethanol and were stored at  $-80^{\circ}\text{C}$ . Samples were purified by centrifugation ( $13,000\times g$  for 30 min at  $4^{\circ}\text{C}$ ), followed by 2 washes with 0.5 mL ice-cold 75% ethanol, followed by resuspending the dried DNA pellet in molecular-grade water. Commercial and breeding program samples were extracted from frozen tissue using the Monarch Genomic DNA Purification Kit (NEB), as per manufacturer's instructions using the enzymatic cleanup protocol with both proteinase K and RNase A, and eluting into 50  $\mu\text{L}$  molecular-grade water.

All samples were quantified by Qubit dsDNA-BR measured by BioSpectrometer (Eppendorf) with a reusable  $\mu\text{Cuvette}$  (G.1.0; Eppendorf). Any sample under 20 ng/ $\mu\text{L}$  was concentrated using the Genomic DNA Concentrator Columns (Zymo Research), eluted in 12  $\mu\text{L}$  water, and quantified by Qubit. Quality was inspected in representative samples from each collection by 1% agarose gel electrophoresis. Samples were normalized to 20 ng/ $\mu\text{L}$  in a 10  $\mu\text{L}$  volume, randomized in 96-well microtiter plates, and submitted to the Institute de Biologie Intégrative et des Systèmes (IBIS) at Université Laval for library preparation. Each plate contained a negative control well. Five samples with concentrations between 17.7 and 20 ng/ $\mu\text{L}$  were also included in a 10  $\mu\text{L}$  volume (i.e. JPN samples 018, 031, 039, 051, and 082).

### ddRAD sequencing library preparation

Selection of restriction enzymes and estimation of required sequencing depth was conducted using SimRAD (Lepais and Weir 2014) using the Yesso scallop reference genome (GCA\_002113885.2; Wang et al. 2017). The estimated haploid genome size for this species is 1.65 Gb (Anisimova 2007; Gregory 2022), but the total sequence length of the assembly is 0.99 Gb, suggesting that the assembly represents 59.8% of the expected genome size. This value was used as an expansion factor for the true expected number of fragments estimated by SimRAD. Enzyme options tested were PstI and MspI, or NsiI and MspI, size selected to retain fragments from 100 to 250 bp. Digestions with NsiI and MspI were predicted in silico to generate 118,850 fragments in the 1.65 Gb genome, whereas PstI and MspI were predicted to have 50,836 fragments; PstI and MspI were selected for digestion.

Samples were multiplexed to 32 individuals per chip, with 6 sequencing chips used. Library preparation was conducted using the semiconductor platform-adapted ddRAD-seq approach (Mascher et al. 2013) with the additional barcodes and size selection step (Abed et al. 2019), as previously described (Sutherland et

**Table 1.** Overview of genotyped samples including sources, sample size per population (*n*), total filtered and demultiplexed reads (millions; M), average  $\pm$  SD reads and alignments per population, and samples passing all filters.

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

al. 2020), but with alternate enzymes. Multiplexed and size-selected libraries were then sequenced on an Ion Torrent Proton at IBIS using the Ion PI Chip Kit v3 chip (Thermo Fisher), as per manufacturer's instructions. Base calling was conducted with the Torrent Suite software (Thermo Fisher), and multiplexed fastq output files were exported using the *FileExporter* plugin.

### Sequence data preprocessing, alignment, and genotyping

Sequence data processing and genotyping largely followed the *stacks\_workflow* repository (E. Normandeau), and all instructions for the analysis is provided (*ms\_scallop\_popgen*; see *Data availability*). Sample barcodes and populations were designated in the metadata file (see *Data availability*). Raw sequence data were quality checked using FastQC (v.0.11.4; Andrews 2010) and MultiQC (v.1.14; Ewels et al. 2016). Sequences were trimmed using cutadapt (v.1.9.1; Martin 2011) to remove too short reads and adapters and inspected again with FastQC and MultiQC. The metadata file was used to demultiplex using the *process\_radtags* module of Stacks v2 with flags *pstI* and *mSP* and truncating all reads to 80 bp (v.2.62; Rochette et al. 2019) in parallel (Tange 2022).

Demultiplexed samples were aligned against the reference genome (ASM211388v2; Wang et al. 2017) using *bwa mem* (Li 2013) then converted to bam format and sorted using *samtools* (Danecek et al. 2021). Read and alignment counts per individual were tallied using custom code (*ms\_scallop\_popgen*; see *Data availability*). Samples with fewer than 500,000 reads were removed including the 3 negative control samples that had between 338 and 2,101 reads each. The *gstacks* module of Stacks was then used with the reference genome to genotype the samples using default settings (*marukilow*, *var\_alpha*: 0.01; *gt\_alpha*: 0.05). Following genotyping, the populations module of Stacks was used to filter, retaining loci genotyped in  $\geq 70\%$  of the individuals per population in all 3 populations (flags: *-r* 0.7, *-p* 3). A global minor allele frequency (MAF) filter retained loci with  $MAF \geq 0.01$  (flag: *-min\_maf* 0.01). To determine the impact of unequal sample size on the mean nucleotide diversity estimated per population, *gstacks* and populations were run as above but with a rarefied data set using 37 samples per population for all 3 populations. Using the full data set, the per-individual inbreeding coefficient was calculated using *vcftools* (*-het*; Danecek et al. 2011), and any outlier individuals with excessive heterozygosity indicating potential genotyping issues were removed. Once outliers were removed, genotyping and filtering were redone with only the retained individuals, as described above.

Multilocus genotypes were exported in VCF and plink formats, where only one of the variants per RAD-tag was retained using the *--write-single-snp* flag of the populations module to produce the single-SNP per locus data set. Furthermore, microhaplotype genotypes were exported in RADpainter and genepop formats. Plink

data were converted from .ped and .map format to .raw format using *plink* (flags: *--recode A --allow-extra-chr*; v.1.90b6.26; Purcell 2022; Purcell et al. 2007). The plink output was then used as input for population genetic analysis in R (R Core Team 2023).

### Population genetic analysis

Single-SNP per locus genotypes were read into R using the *read.PLINK* function of *adegenet* (Jombart and Ahmed 2011). Data were formatted and converted to *genind* format using the *df2genind* function of *adegenet*. Functions applied were used from the *simple\_pop\_stats* repository (see *Data availability*). Genotyping rate was calculated per sample using the *percent\_missing\_by\_ind* function of *simple\_pop\_stats*, and individuals were retained if they had less than 30% missing data. MAF was calculated per locus, and any locus with  $MAF < 0.01$  was removed. Per locus  $F_{ST}$  was calculated using *pegas* (Paradis 2010).  $H_{OBS}$  was calculated using *adegenet*, and a test of Hardy-Weinberg (HW) equilibrium was conducted per locus using *pegas*. Loci showing significant deviation from HW proportions ( $P \leq 0.01$ ) in any 1 of 3 populations were identified and removed from the data set. Loci with global  $H_{OBS} \geq 0.5$  were also removed.

Private alleles per population were identified using the *private\_alleles* function of *poppr* (Kamvar et al. 2014) using the filtered data set. Region-specific private alleles were also identified in the filtered data set by combining all Canadian samples and comparing to the JPN samples. MAF distributions were compared among populations, including all loci or loci that included private alleles. Per locus  $H_{OBS}$  was also calculated for each population specifically to identify loci with high heterozygosity in both JPN and VIU populations and plotted using *ggplot2* (Wickham 2016).  $N_e$  was calculated using *NeEstimator* v.2.0 (Do et al. 2014) using both single-SNP per locus data and microhaplotype data. Interindividual relatedness was calculated using single-SNP per locus data using *related* (Pew et al. 2015) and generally only considered comparisons within a population. Relatedness was also calculated using microhaplotype data using *fineRADstructure* (Malinsky et al. 2018). Outlier thresholds for pairwise relatedness values generated from the single-SNP per locus data were identified from boxplots, focusing on the outlier level for the VIU population. Once a threshold was determined, a purged close relative data set was created to reduce the impact of putative parents or sibs on downstream PCA or  $F_{ST}$  calculations by removing 1 individual from the pair until no pairs remained above the set cutoff. The purged close relative data set was refiltered for low MAF variants and then used as an input to a principal component analysis (PCA) using the *glPca* function of *adegenet*. The purged close relative data set was then used to calculate mean population-level  $F_{ST}$  that was calculated using 1,000 bootstraps to generate 95%

**Table 2.** Genotyping summary statistics and analysis results for both the full and rarefied data sets, including the number of polymorphic sites and polymorphism rate, and  $H_{OBS}$  based on all sites (polymorphic or monomorphic).

Dataset	Pop	Samples/locus	Poly. sites	Poly. rate (%)	$H_{OBS}$	$N_e$ (95% C.I.)	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	5,423	1,401
	VIU	63.7	15,471	1.35	0.00276	26.4–26.4	4,531	270
Rarefied	BC	34.2	12,685	0.73	0.00186	9.5–9.5	—	—
	JPN	35.7	17,490	1.00	0.00177	$\infty$	—	—
	VIU	35.4	14,765	0.85	0.00184	22.9–23.0	—	—

Both data sets allow multiple variants per RAD-tag; the full data set has 21,048 variants in 1,143,426 sites, and the rarefied had 20,269 variants in 1,746,980 genomic sites.  $N_e$  was calculated based on microhaplotypes (9,377 loci). The number of filtered loci shows the number of polymorphic loci per population in single-SNP per locus data after filters in each population including per population MAF and HW ( $P < 0.01$ ).

confidence limits using the boot.ppfst function of hierfstat (Goudet and Jombart 2022).

## Results

### Sequencing, genotyping, and filtering

In total, 558,098,638 single-end reads were generated from 189 Yesso scallops sequenced on 6 Ion Torrent chips. Each chip produced on average ( $\pm$ SD) 93.0  $\pm$  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 ( $\pm$ SD) number of reads was 2.7  $\pm$  1.2 M (Table 1). Samples with the fewest reads were from the JPN marine surveys with an average number of reads per sample of 1.7 M ( $\pm$ 0.6 M), and those with the most were from the BC farm samples with an average number of reads per sample of 3.5 M ( $\pm$ 0.9 M). One JPN marine survey scallop (JPN\_075) was removed due to low coverage, and 1 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 to 81.6% (Supplementary 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  $\pm$  28.7 $\times$  (range: 2.9–176.6 $\times$ ). Filtering to remove loci with excess missing data (see Materials and methods) or low global MAF (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, 2, or 3 variants were observed in 3,721 (39.7%), 2,585 (27.6%), and 1,541 (16.4%) tags, respectively (Supplementary 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. Considering RAD-tags as microhaplotypes identified counts of 2, 3, or 4 alleles per tag in 74% of the tags (i.e. 4,073, 2,609, and 1,431 RAD-tags, respectively; Supplementary File S3). Five to nine alleles per tag were observed in 1,215 tags (23.8%), 45 tags had between 10 and 16 alleles, and 4 had over 20 alleles each. Notably, these tallies include variants or microhaplotypes not yet filtered for deviations from HW proportions.

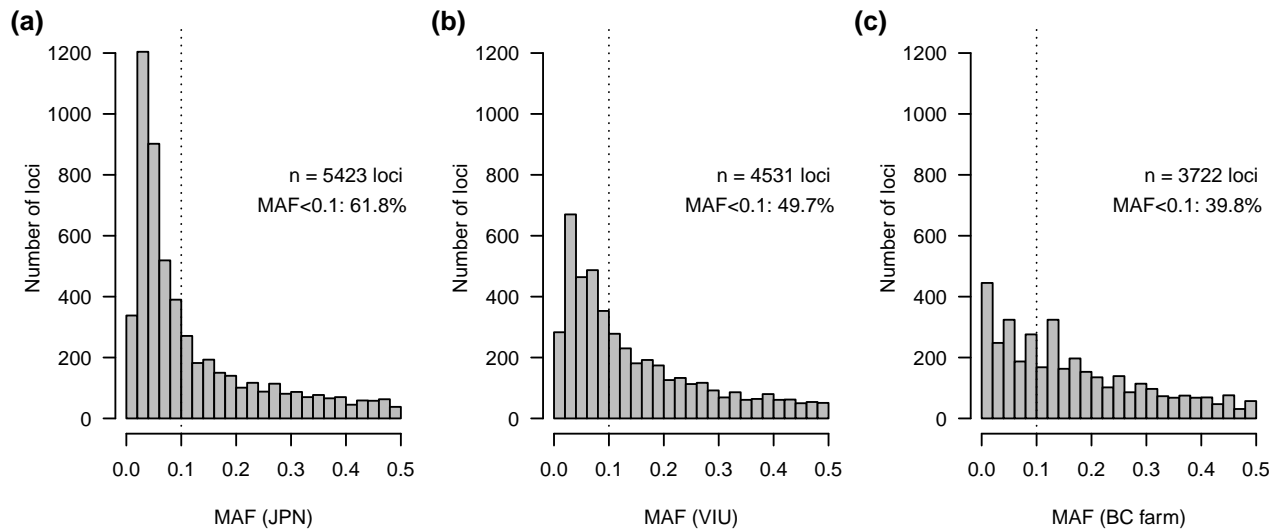
Per individual missing data was on average 8.1  $\pm$  12.2%. Samples with more than 30% missing data were dropped from the analysis, which removed 2 samples from VIU and 14 samples from JPN (see Table 1 for sample size post filters). After filtering on missing data, 173 individuals remained, with an average per sample missing data of 5.0  $\pm$  4.3%. Of the 9,375 variants in the single variant per RAD-tag data set, 2,578 (27.5%) did not conform to expected HW proportions ( $P < 0.01$ ) in at least 1 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 data set had 6,547 variants remaining for downstream analyses.

### Genetic diversity, low-frequency variants, and private alleles

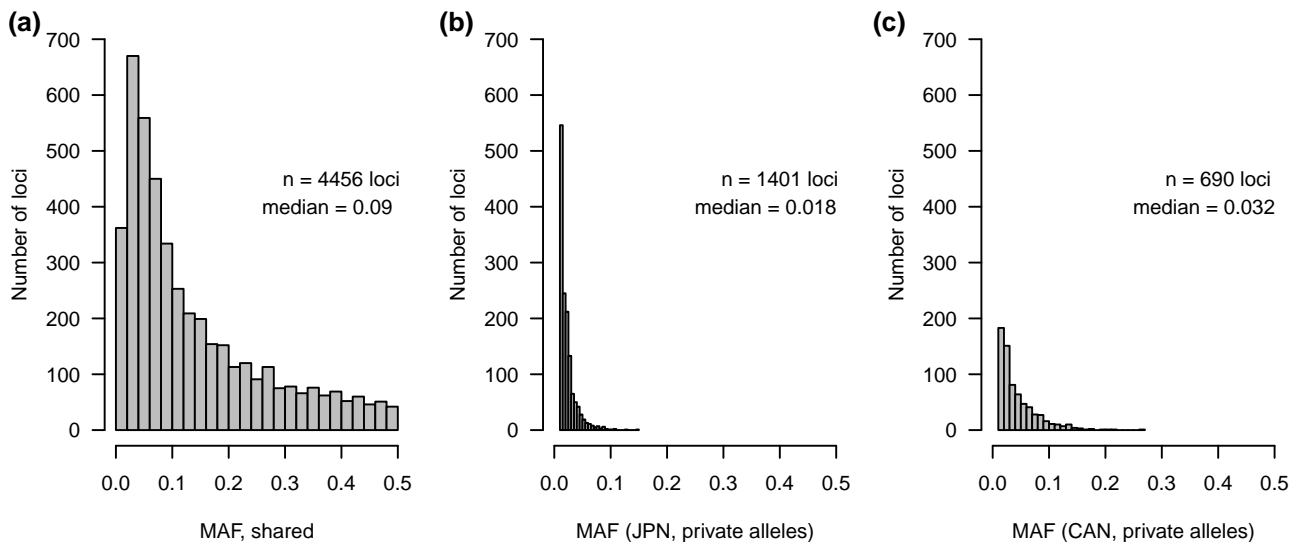
Considering all 21,048 variants (i.e. multiple SNPs per RAD-tag) prior to HW filters, the JPN collection had the highest polymorphism rate at 1.7%, with 19,411 variants (Table 2). By comparison, there were 15,471 variants in the VIU collection (i.e. 1.35% polymorphism rate) and 12,286 variants in the BC farm collection (i.e. 1.07% polymorphism rate). However,  $H_{OBS}$  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 (SE for all:  $\pm$ 0.00003). Using the rarefied data set ( $n = 37$  samples per population), these relative trends in polymorphism rate and  $H_{OBS}$  remained consistent (Table 2). 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 data set, 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 data sets retained 5,423 (82.8%), 4,531 (69.2%), and 3,722 (56.9%) variants, respectively (Fig. 1). In addition to having the most variants retained, JPN also had the highest proportion of retained lower frequency variants (i.e. MAF between 0.01 and 0.10) with 61.8% of the variants in this range, followed by VIU (49.7%), and lastly the BC farm (39.8%).

Private alleles within the single-SNP per locus filtered data set were also most numerous in JPN with 1,401 private alleles, followed by VIU ( $n = 270$ ) and BC farm ( $n = 58$ ). Regionally, when VIU and BC farm were merged into a Canadian population then compared to JPN, there were 1,401 private alleles in JPN and 690



**Fig. 1.** Per locus MAF distribution for the JPN wild population a), the VIU breeding population b), and the BC farm population c) following all filters, including population-specific low MAF filters.



**Fig. 2.** MAF distributions for a) loci found in all populations, b) private alleles in JPN, and c) private alleles in Canada (VIU and BC farm).

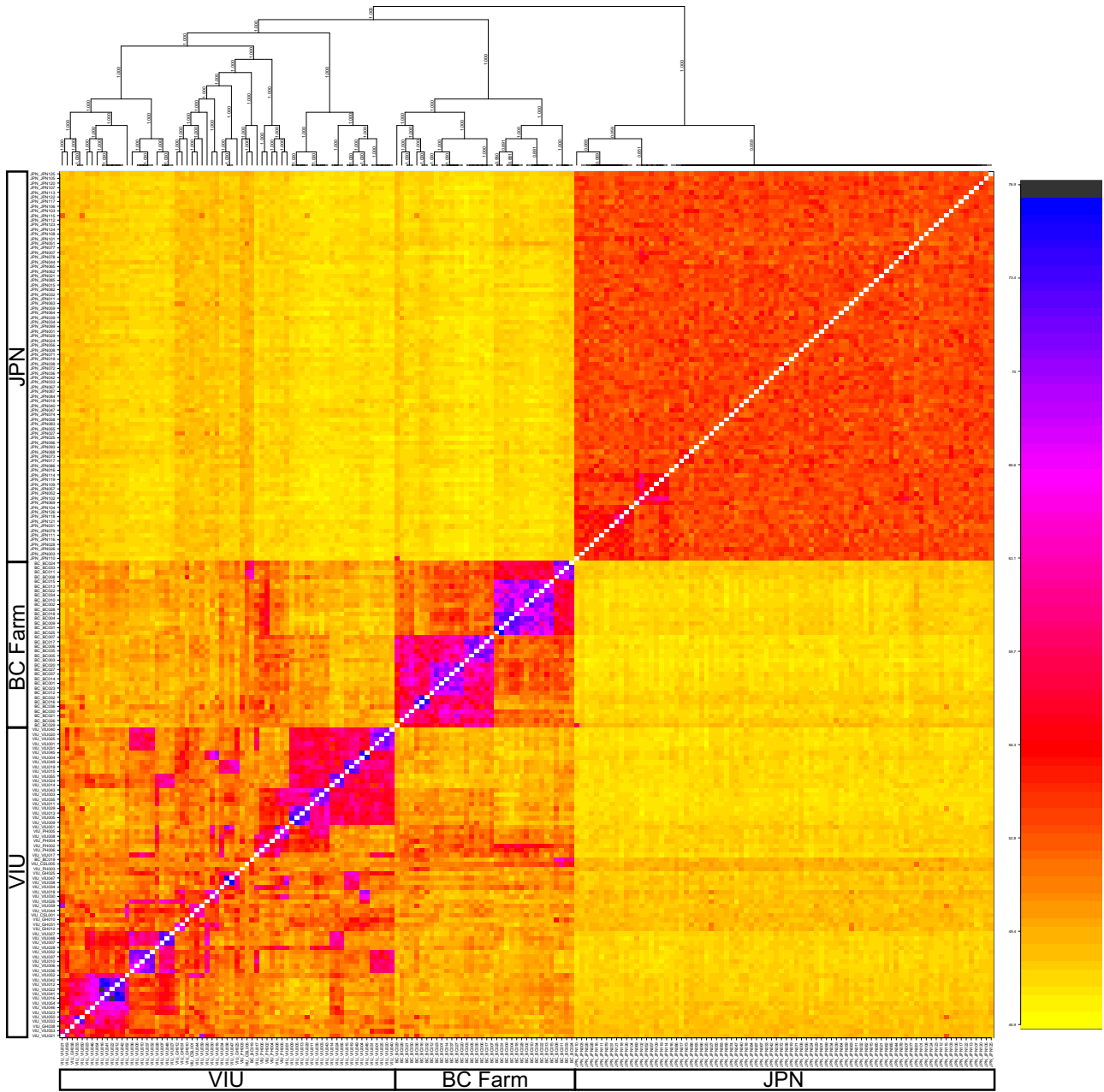
in Canada. Although less numerous than JPN, the Canadian private alleles were higher overall in MAF (median MAF = 0.032) than the private alleles in JPN (median MAF = 0.018; Fig. 2).

$N_e$  was estimated for each population using either the microhaplotype multiple-SNP per RAD-tag data set ( $n = 21,048$  variants, 9,377 RAD-tags) or the single-SNP per RAD-tag data set ( $n = 9,375$  variants). The linkage disequilibrium (LD) method estimated 95% confidence interval (C.I.) of  $N_e$  for JPN as 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 9.3–9.3 ( $n = 6,839$  tags) when using a  $P_{\text{CRIT}}$  value of 0.01 (Table 2). The heterozygosity excess method was unable to estimate the 95% C.I. for variants with MAF > 0.01, and the molecular coancestry method estimates were very low for all populations (estimated  $N_{\text{eb}}$  2.6–6.5; see Supplementary File S4). Using the single-SNP per RAD-tag data, the 95% C.I. estimates of  $N_e$  with the LD method were not calculable using the data for JPN (lower and upper values were both infinite) but were estimated at similar levels for VIU 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 rarefied data set ( $n = 37$  individuals per population), 95% C.I. of  $N_e$  for VIU and BC farm were again estimated to be 9.5 and 22.9–23.0, respectively, but the JPN population  $N_e$  was not estimable with the existing data (Supplementary File S4).

### Interindividual relatedness

Using the microhaplotype data in fineRADstructure (Malinsky et al. 2018), clusters of individuals with elevated relatedness were observed in the VIU and BC farm collections, whereas the JPN collection was comprised of approximately equally dissimilar samples (Fig. 3). The VIU collection has many small clusters of related individuals, and some of these individual clusters are estimated to be more different from each other than are the clusters in the BC farm, based on shared microhaplotypes. Within the JPN collection, slight clustering is observed in a grouping of 7 individuals and a grouping of 11 individuals. The cause of these slight groupings is not clear given that it encompasses both marine



**Fig. 3.** Similarity between pairs of individuals based on shared microhaplotypes. Unsupervised clustering positioned all samples within their respective groupings (i.e. VIU, BC farm, and JPN) except for 1 BC farm sample that clustered within the VIU grouping (BC019). The proportion of shared microhaplotypes are shown by the color scale bar, where blue/black (darkest colour) is the highest level of shared microhaplotypes and yellow (lightest colour) is the lowest. The dendrogram (top) shows groupings of individuals based on genetic similarity, with clusters of individuals observed within breeding program groupings.

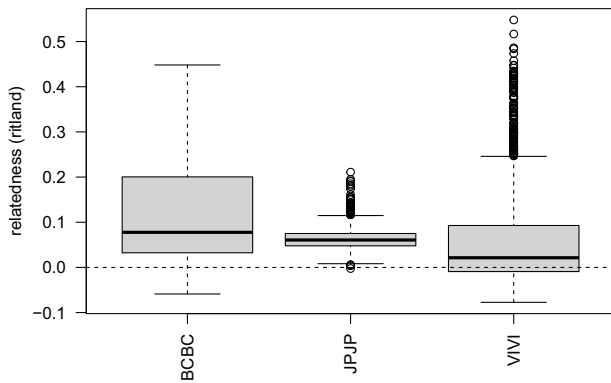
survey and commercial harvest samples. When considering relatedness using the single-SNP per locus data set, the JPN collection is observed to have a relatively tight distribution of relatedness compared to the VIU or BC farm collections (Fig. 4). Many highly related outliers are observed in the VIU collection, as expected given the groups of related individuals in the microhaplotype analysis for this collection. One JPN sample (JPN\_110) showed high relatedness to BC\_029, and the cause of this is unknown, given that each sample also shows high similarity to other members of their own collections.

To reduce the impacts of family structure on downstream PCA (Patterson et al. 2006) and pairwise population-level  $F_{ST}$  contrasts, outlier pairs were determined to be at Ritland relatedness

0.25 for the VIU population (Fig. 4), and so an individual from each pair above this cutoff was removed from these putatively close relatives (see Materials and methods). This resulted in the removal of 43 of 65 (66%) of the VIU samples and 28 of 37 (76%) of the BC samples. None of the JPN samples were removed, and so VIU, BC, and JPN populations had 22, 9, and 71 samples each for PCA and  $F_{ST}$  analyses (see below). After the putative close relatives were removed, an additional MAF filter was applied, removing any variants with  $MAF < 0.01$ , which resulted in the removal of 283 additional variants, leaving 6,264 variants in the data set. Per locus  $F_{ST}$  was also recalculated, and both the purged close relatives and all sample data sets are presented in Supplementary File S5.

## Global genetic differentiation

In an unsupervised PCA using the single-SNP per locus data with putative close relatives removed ( $n=6,264$  variants), samples were separated by country (i.e. Canada or JPN) across PC1, explaining 5.7% of the overall variation (Fig. 5a). The BC farm and VIU samples were spread across PC2 (2.5% of the variance explained), with overlapping 95% C.I. ellipses, and these axes captured within-farm variation (Fig. 5b). After PC2, individual PCs were less informative as indicated by the scree plot (Fig. 5a inset). Analysis with putative relatives included resulted in a greater separation between the farms across PC2, as expected with family structure within the collections (Patterson et al. 2006; Elhaik 2022; Supplementary File S6).



**Fig. 4.** Interindividual relatedness values for all pairs of individuals within each population as evaluated by the Ritland statistic, using a single-SNP per locus. BCBC, BC farm; JPJP, Japan; VIVI, VIU.

Genetic differentiation analysis on the purged putative sibs data set using  $F_{ST}$  (Weir and Cockerham 1984) indicates significant global genetic differences between all of the populations (Table 3). Similar genetic differentiation was observed between the VIU hatchery and the JPN wild population (95% C.I.  $F_{ST}=0.061-0.068$ ) and the BC farm and JPN wild ( $F_{ST}=0.068-0.077$ ). The 2 cultured populations had slightly lower differentiation from each other than either with the wild JPN samples but still relatively high differentiation ( $F_{ST}=0.046-0.055$ ). As observed in the PCA, the removal of putative close relatives reduced  $F_{ST}$  differences between the hatcheries (Supplementary File S6).

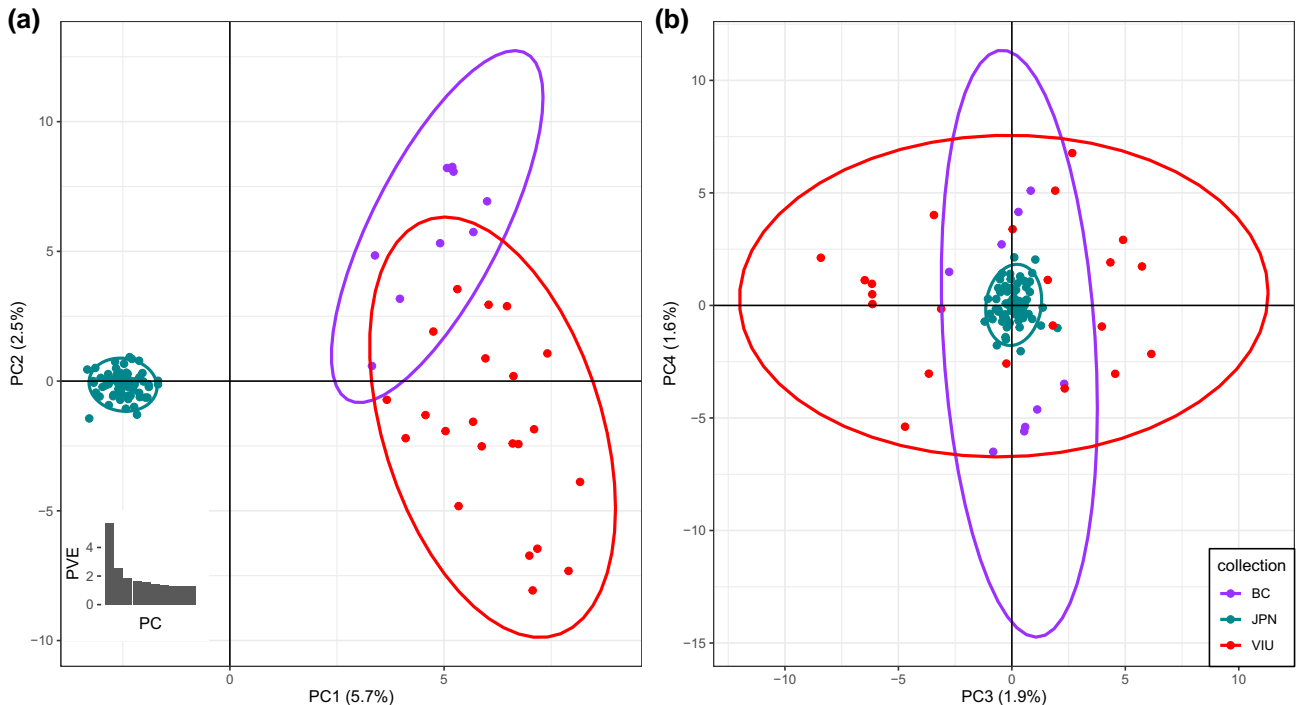
## Per locus heterozygosity and highly variable tags

Filtered single-SNP per RAD-tag markers were inspected for per-locus  $H_{OBS}$  within individual populations of JPN or VIU, and then values were compared to identify variants that are expected to have high heterozygosity in both collections (Fig. 6; Supplementary File S5 for per-locus values). Additionally, given the observation of RAD-tags with high numbers of variants per tag, the marker names, the number of variants per tag, and whether the tag is within HW proportions are shown in

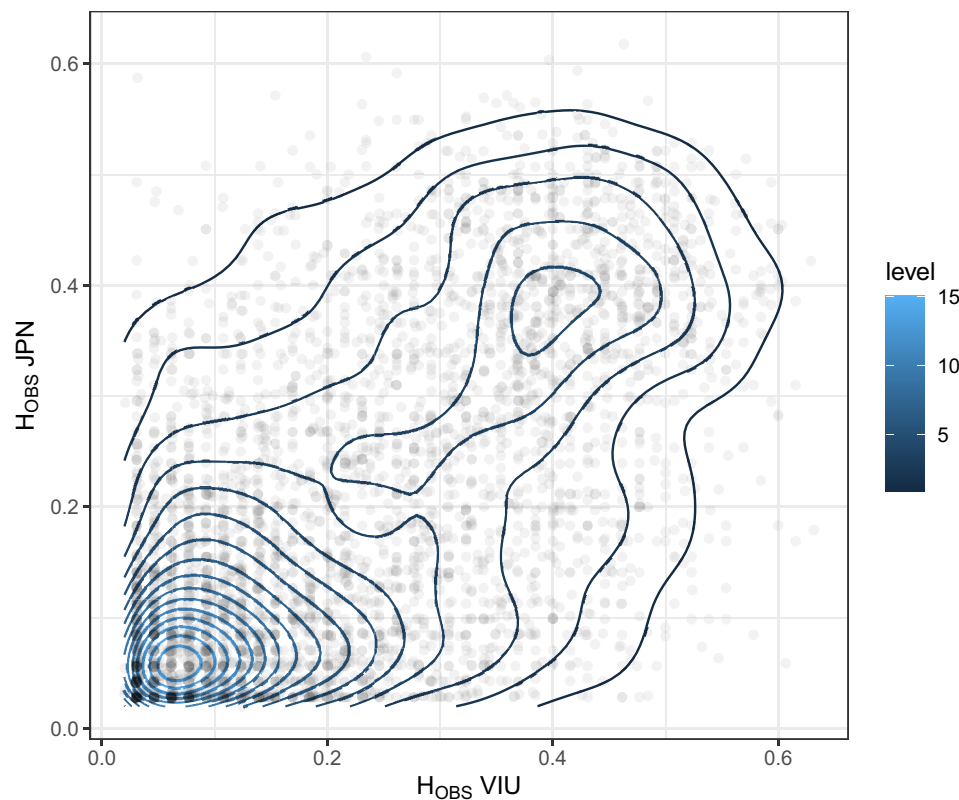
**Table 3.** Population genetic differentiation ( $F_{ST}$ ) comparison expressed as 95% C.I. with the upper limit on the upper right and lower limit in the lower left.

	BC	JPN	VIU
BC	—	0.077	0.055
JPN	0.068	—	0.068
VIU	0.046	0.061	—

Calculations used the single-SNP per locus data after filters on the data set with close relatives removed.



**Fig. 5.** PCA showing clustering of the 3 collections. a) Farmed or breeding program scallops in BC, Canada, are separated from wild scallops in JPN across PC1 (5.7% variance explained). The 2 BC collections are spread across PC2, but the 95% C.I. ellipses are overlapped, and a lower proportion of variance is explained by this axis (2.5%). b) PC3 and PC4 separate variation within BC collections. After the first 2 PCs, the percent variation explained reduces as shown in the scree plot (see inset in a).



**Fig. 6.** JPN and VIU population-specific  $H_{OBS}$  comparison using the filtered, single-SNP per locus data set. Note: the per locus heterozygosity values were generated using population-specific  $H_{OBS}$  values in the data set that still contained putative close relatives.

**Supplementary File S5.** The single-SNP per locus and microhaplotype data set VCF files are provided, which give the location of the markers in the reference genome (see *Data availability*).

## Discussion

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

### Genomic diversity in wild Yesso scallop from northern JPN

Wild scallops from northern JPN ( $n = 71$ ) were estimated to have a 1.7% polymorphism rate when including all SNPs with  $MAF > 0.01$  that were present in at least 70% of the samples from each of 3 populations and were calculated prior to removing variants that did not conform to HW proportions. This rate was estimated based on 0.08% (or 1.14 Mb) of the expected 1.43 Gb genome (Wang et al. 2017) and fits with the expectation of high polymorphism rates for shellfish (Plough 2016). The polymorphism rate of the wild, hermaphrodite parent individual used for the Yesso scallop reference genome assembly was estimated at 1.04% (Wang et al. 2017), and the individual used for a Pacific oyster genome assembly was evaluated at 1.30% (Zhang et al. 2012). Furthermore, the king

scallop *Pecten maximus* sequenced individual was estimated to have a 1.7% heterozygosity rate, a higher rate than the Pacific oyster assembly individual, which the authors ascribe to the breeding program source of the sequenced Pacific oyster (Kenny et al. 2020). Sutherland et al. (2020) in an analysis of a variety of wild, naturalized, and farmed Pacific oysters identified a mean overall population-level polymorphism rate of 1.43% with lower rates in farmed populations (e.g. 0.94–1.13%). The high polymorphism rate in Pacific oyster, also estimated at 1 SNP per 40 bp (2.5%), was given as one of the reasons for pursuing the reference genome, as well as a challenge to overcome in order to do so (Hedgecock et al. 2005). Combining the high polymorphism rate with SRS effects on annual variation in allele frequencies leads to the question of what the longevity is of many of the variants identified here in the lower frequency range (i.e. 62% of variants with  $MAF$  between 0.01 and 0.10).

The high polymorphism rate resulted in some RAD-tags having numerous variants within the same 80 bp fragment, where 16% of tags had more than 4 variants. These highly variable regions may need to be treated with caution, since the genome assembly may contain collapsed segments, given the differences between the expected and the assembled genome size (Wang et al. 2017). High repeat content resulting in missing genomic segments was also an issue for the Pacific oyster (Zhang et al. 2012), which is thought to be comprised of 35% repetitive elements (Hedgecock et al. 2005).

The  $N_e$  of the wild JPN collection characterized here was estimated to be between 25,048 and 56,291 breeding individuals. The census size of the population around Hokkaido is not known, but approximately 145,000–375,000 individuals are expected to have been harvested annually in the bay from which samples were collected (Uchiura Bay, Itoh N, personal communication).



## Cultured strains and genomic impacts of breeding programs

The BC cultured scallop populations characterized here have significantly lower polymorphism rates than the source population. The genotyped VIU collection, which likely represents the full complement of genetic backgrounds in the VIU breeding program (Green T, personal observation), was estimated to have a polymorphism rate of 1.35%, which is lower than the source wild population (i.e. 1.7%). The BC farm, with an estimated polymorphism rate of 1.07%, is likely only a subset of the full genetic variation in the cultured lineages from the seed provider for the farm since this collection was obtained as farmed adults from a single sampling event. Additionally, the JPN collections are obtained from several years of sampling, and so this does encompass a larger number of spawning events. Notably, both cultured collections were also typified by a depletion of alleles in the lower MAF range (i.e. MAF 0.01–0.10).

Genetic changes from source populations are also observed as skewed allele frequencies genome-wide, which can occur through founder effects and drift. For example, Suminoe oyster hatchery strains were differentiated both against the source population and between each strain ( $F_{ST} = 0.05$ – $0.24$ ; Xiao et al. 2011). These impacts are likely due to few parents being included in the original founding population (i.e. founder effects) and drift occurring due to a low  $N_e$  but potentially also from any artificial selection that has occurred (Li et al. 2007; Xiao et al. 2011). Hatchery strains of the common blue mussel *Mytilus edulis* were found to be significantly different from each other ( $F_{ST} = 0.03$ – $0.08$ ) as well as from the source wild population ( $F_{ST} = 0.07$ – $0.08$ ), which accompanies reduced  $N_e$  and allelic diversity (Gurney-Smith et al. 2017). Eastern oyster hatchery strains also show significant differentiation ( $F_{ST} = 0.076$ ) between hatchery and wild populations (Carlsson et al. 2006), and eastern oyster strains increase in their differentiation within and between hatchery strains over time (Varney and Wilbur 2020). Pacific oyster hatchery strains showed significant differentiation from source populations, dwarfing the differentiation observed between natural populations in different countries along the same translocation lineage (Sutherland et al. 2020). In Yesso scallop, here we observed high differentiation between hatchery lineages as well as between both hatchery lineages with the source population ( $F_{ST}$ : 0.05–0.08). As a notable contrast, 5 different farms characterized in China showed generally low differentiation between farms ( $F_{ST} = 0.02$ – $0.03$ ) but high differentiation between the farms and the source populations ( $F_{ST} = 0.09$ – $0.15$ ; Li et al. 2007). It is possible that this similarity between the farms in China may indicate a common seed scallop provider for the different farms. In general, the cultured lineages here show the expected level of differentiation from the wild population relative to these other shellfish species, many of which were referenced as a concern in the original studies and needing to be monitored for continued diversity losses.

The  $N_e$  of the cultured lineages here was significantly reduced relative to the wild population, indicating founder effects and drift. The  $N_e$  estimated for the VIU population was 26, conforming to the expectations based on the known number and diversity of the source parents used to initiate this breeding population (Green T, personal observation). The BC farm population, which may not fully represent the entire broodstock, as described above, was estimated at  $N_e = 9$ . These values are remarkably lower than that observed for the wild population (i.e.  $N_e = 25,048$ – $56,291$ ), which required all data to estimate and was only estimable using

microhaplotype information. Similarly, cultured stocks of Yesso scallop analyzed by Li et al. (2007) had  $N_e$  calculated between 27 and 70, even though hundreds of males and females are used to propagate the hatchery strains. These levels were a concern regarding the long-term sustainability of the stock (Li et al. 2007). Similarly, Gurney-Smith et al. (2017) estimated  $N_e$  values under 50 for the 3 hatchery strains of common blue mussel analyzed, and the wild population was too large to estimate with the available data. Considering the genome-wide scope of the allele frequency changes, and the significant decreases in  $N_e$ , it is likely that founder effects and drift are sufficient to explain the allele frequency shifts observed in the present study.

An additional indicator of founder effects and drift is that private alleles in the cultured lineages were of relatively high frequency (MAF = 0.01–0.28) relative to the wild population. Although these private alleles are likely to be in the JPN wild population, they were not represented in the sampling here and therefore not expected at the same high frequency as those in culture. In contrast, wild private alleles, although more numerous, were lower in frequency (MAF = 0.01–0.15).

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

Breeding programs in general are challenged by losses of genetic variation over successive generations of breeding with or without selection, which in shellfish is further compounded by high fecundity and large differences in reproductive success, resulting in significant drift (Boudry et al. 2002; Varney and Wilbur 2020). This problem becomes even more challenging if the source population is not available for replenishing the broodstock through introgression, as is the case when the cultured stock is grown outside of its native range (Li et al. 2007; Gurney-Smith et al. 2017). This is the situation for Yesso scallop in BC, where no naturalized populations exist outside of hatchery breeding programs. This increases the need for effective management of shellfish hatcheries

including genetic diversity monitoring (Gurney-Smith et al. 2017), good record keeping, and the availability of genetic tools to confirm pedigrees, low relatedness of parents in target crosses, and confirmed identity of lineage for the individuals chosen as broodstock for crosses (Hedgecock and Davis 2007).

## Conclusion

The wild Yesso scallop characterized from JPN has a high polymorphism rate that is 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  $N_e$  around 9–26 individuals relative to that calculated from the wild population at 25,048–56,291 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  $H_{OBS}$  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  $H_{OBS}$  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 and 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](https://github.com/bensutherland/ms_scallop_popgen). Analysis pipeline applied (genotyping): [https://github.com/enormandeaustacks\\_workflow](https://github.com/enormandeaustacks_workflow). Analysis pipeline applied (analysis): [https://github.com/bensutherland/simple\\_pop\\_stats](https://github.com/bensutherland/simple_pop_stats). Raw sequencing data have been uploaded to SRA under BioProject PRJNA947158 and 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.

Supplemental material available at G3 online.

## Acknowledgments

The authors thank the editor and an anonymous reviewer for their comments on an earlier version of the manuscript.

## Funding

This work was supported by the Province of British Columbia, Ministry of Agriculture and Food and an anonymous donor.

## Conflicts of interest

BJGS is affiliated with Sutherland Bioinformatics. The author has no competing financial interests to declare. The other authors declare that no competing interests exist.

## Literature cited

- Abed A, Légaré G, Pomerleau S, St-Cyr J, Boyle B, Belzile FJ. 2019. Genotyping-by-sequencing on the Ion Torrent platform in barley. *Methods Mol Biol.* 1900:233–252. doi:10.1007/978-1-4939-8944-7\_15.
- 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. *Comp Cytogenet.* 1(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(3–4):283–296. doi:10.1016/S0044-8486(01)00841-9.
- Bourne NF. 2000. The potential for scallop culture—the next millennium. *Aquac Int.* 8(2/3):113–122. doi:10.1023/A:1009212226803.
- Bower SM, Blackburn J, Meyer GR, Welch DW. 1999. Effect of *Perkinsus qugwadi* on various species and strains of scallops. *Dis Aquat Organ.* 36(2):143–151. doi:10.3354/dao036143.
- Carlsson J, Morrison CL, Reece KS. 2006. Wild and aquaculture populations of the eastern oyster compared using microsatellites. *J Hered.* 97(6):595–598. doi:10.1093/jhered/esl034.
- Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, Handsaker RE, Lunter G, Marth GT, Sherry ST, et al. 2011. The variant call format and VCFtools. *Bioinformatics* 27(15):2156–2158. doi:10.1093/bioinformatics/btr330.
- Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, Pollard MO, Whitwham A, Keane T, McCarthy SA, Davies RM, et al. 2021. Twelve years of SAMtools and BCFtools. *GigaScience* 10(2):2. doi:10.1093/gigascience/giab008.
- 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 ( $N_e$ ) from genetic data. *Mol Ecol Resour.* 14(1):209–214. doi:10.1111/1755-0998.12157.
- Elhaik E. 2022. Principal component analyses (PCA)-based findings in population genetic studies are highly biased and must be reevaluated. *Sci Rep.* 12(1):14683. doi:10.1038/s41598-022-14395-4.
- Evans F, Matson S, Brake J, Langdon C. 2004. The effects of inbreeding on performance traits of adult Pacific oysters (*Crassostrea gigas*). *Aquaculture* 230(1–4):89–98. doi:10.1016/j.aquaculture.2003.09.023.
- Ewels P, Magnusson M, Lundin S, Käller M. 2016. MultiQC: summarize analysis results for multiple tools and samples in a single report. *Bioinformatics* 32(19):3047–3048. doi:10.1093/bioinformatics/btw354.
- 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. doi:10.1016/j.aquaculture.2016.08.038.
- Hedgecock D. 1994. Does variance in reproductive success limit effective population sizes of marine organisms. In: Beaumont AR, editor. *Genetics and evolution of aquatic organisms*. London (UK) : Chapman and Hall. p. 122–134.
- Hedgecock D, Davis JP. 2007. Heterosis for yield and crossbreeding of the Pacific oyster *Crassostrea gigas*. *Aquaculture* 272:S17–S29. doi:10.1016/j.aquaculture.2007.07.226.
- Hedgecock D, Gaffney PM, Gouilletquer P, Guo X, Reece K, Warr GW. 2005. The case for sequencing the Pacific oyster genome. *J Shellfish Res.* 24(2):429–441. 413. doi:10.2983/0730-8000.

- Hedgecock D, McGoldrick DJ, Bayne BL. 1995. Hybrid vigor in Pacific oysters: an experimental approach using crosses among inbred lines. *Aquaculture* 137(1–4):285–298. doi:10.1016/0044-8486(95)01105-6.
- Hedgecock D, Pudovkin AI. 2011. Sweepstakes reproductive success in highly fecund marine fish and shellfish: a review and commentary. *Bull Marine Sci.* 87(4):971–1002. doi:10.5343/bms.2010.1051.
- Hedgecock D, Sly F. 1990. Genetic drift and effective population sizes of hatchery-propagated stocks of the Pacific oyster, *Crassostrea gigas*. *Aquaculture* 88(1):21–38. doi:10.1016/0044-8486(90)90316-F.
- Hedrick PW. 2005. Inbreeding, pp. 237–292 in *Genetics of Populations*. Sudbury (MA): Jones and Bartlett Publishers.
- Holden JJ, Collicutt B, Covernton G, Cox KD, Lancaster D, et al. 2019. Synergies on the coast: challenges facing shellfish aquaculture development on the central and north coast of British Columbia. *Mar Policy*. 101:108–117. doi:10.1016/j.marpol.2019.01.001.
- Jombart T, Ahmed I. 2011. ADEGENET 1.3–1: new tools for the analysis of genome-wide SNP data. *Bioinformatics* 27(21):3070–3071. doi:10.1093/bioinformatics/btr521.
- 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. doi:10.7717/peerj.281.
- Kenny NJ, McCarthy SA, Dudchenko O, James K, Betteridge E, Corton C, Dolucan J, Mead D, Oliver K, Omer AD, et al. 2020. The gene-rich genome of the scallop *Pecten maximus*. *GigaScience* 9(5):5. doi:10.1093/gigascience/giaa037.
- Kobayashi T, Kijima A. 2010. Effects of inbreeding depression in Pacific abalone *Haliotis discus hannai*. *J Shellfish Res.* 29(3):643–649. doi:10.2983/035.029.0313.
- 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. *Mol Ecol Resour.* 14(6):1314–1321. doi:10.1111/1755-0998.12273.
- Li H. 2013. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM, pp. 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(1–4):211–219. doi:10.1016/j.aquaculture.2007.04.017.
- Malinsky M, Trucchi E, Lawson DJ, Falush D. 2018. RADpainter and fineRADstructure: population inference from RADseq data. *Mol Biol Evol.* 35(5):1284–1290. doi:10.1093/molbev/msy023.
- Martin M. 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet J.* 17(1):3. doi:10.14806/ej.17.1.200.
- 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(10):e76925. doi:10.1371/journal.pone.0076925.
- Meyer GR, Lowe GJ, Gilmore SR, Bower SM. 2017. Disease and mortality among Yesso scallops *Patinopecten yessoensis* putatively caused by infection with *Francisella haliotidica*. *Dis Aquat Organ.* 125(1):79–84. doi:10.3354/dao03130.
- Paradis E. 2010. Pegas: an R package for population genetics with an integrated-modular approach. *Bioinformatics* 26(3):419–420. doi:10.1093/bioinformatics/btp696.
- Patterson N, Price AL, Reich D. 2006. Population structure and eigenanalysis. *PLoS Genet.* 2(12):e190. doi:10.1371/journal.pgen.0020190.
- 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(3):557–561. doi:10.1111/1755-0998.12323.
- Plough LV. 2016. Genetic load in marine animals: a review. *Curr Zool.* 62(6):567–579. doi:10.1093/cz/zow096.
- Purcell SL. 2022. PLINK (v.1.90b6.26).
- Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, Maller J, Sklar P, de Bakker PIW, Daly MJ, et al. 2007. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet.* 81(3):559–575. doi:10.1086/519795.
- 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. *Mol Ecol.* 28(21):4737–4754. doi:10.1111/mec.15253.
- Solé-Cava AM, Thorpe JP. 1991. High levels of genetic variation in natural populations of marine lower invertebrates. *Biol J Linn Soc.* 44(1):65–80. doi:10.1111/j.1095-8312.1991.tb00607.x.
- 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. *Evol Appl.* 13(6):1380–1399. doi:10.1111/eva.12965.
- 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. doi:10.1016/j.aquaculture.2020.735452.
- 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(1–2):59–65. doi:10.1016/0044-8486(94)90282-8.
- Wang S, Zhang J, Jiao W, Li J, Xun X, Sun Y, Guo X, Huan P, Dong B, Zhang L, et al. 2017. Scallop genome provides insights into evolution of bilaterian karyotype and development. *Nat Ecol Evol.* 1(5):0120. doi:10.1038/s41559-017-0120.
- Weir BS, Cockerham CC. 1984. Estimating F-statistics for the analysis of population structure. *Evolution* 38(6):1358–1370. doi:10.1111/j.1558-5646.1984.tb05657.x.
- Wickham H. 2016. ggplot2: Elegant Graphics for Data Analysis. New York: Springer-Verlag.
- Williams GC. 1975. Sex and Evolution. (MPB-8). Princeton University Press, 8.
- 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. *J Shellfish Res.* 30(3):751–760, 710. doi:10.2983/035.030.0315.
- Yu Z, Guo X. 2004. Genetic analysis of selected strains of eastern oyster (*Crassostrea virginica* Gmelin) using AFLP and microsatellite markers. *Marine Biotechnol.* 6(6):575–586. doi:10.1007/s10126-004-3600-5.
- Zhang G, Fang X, Guo X, Li L, Luo R, Xu F, Yang P, Zhang L, Wang X, Qi H, et al. 2012. The oyster genome reveals stress adaptation and complexity of shell formation. *Nature* 490(7418):49–54. doi:10.1038/nature11413.
- 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 Biotechnol.* 19(3):301–309. doi:10.1007/s10126-017-9751-y.
- 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 Biotechnol.* 21(5):623–633. doi:10.1007/s10126-019-09907-9.