Introgression between wild and selectively bred Sydney Rock Oysters, *Saccostrea glomerata*

L. Auckett

18/06/2021

# Abstract Introgressive gene flow can occur when wild and selectively bred populations occur within the same geographic space. The Sydney Rock Oyster *Saccostrea glomerata* is an economically significant bivalve species capable of providing a number of ecosystem services. Wild populations inhabit spaces adjacent to selectively bred and farmed populations of Sydney Rock Oysters. However, Sydney Rock Oysters are functionally extinct in the wild, with the more resilient members of wild populations being selected and removed for use in breeding programs. Here, populations of wild and farmed oysters were collected and analyzed to determine the level of introgressive gene flow occurring. This study shows that introgressive gene flow is not occurring between wild and selectively bred oysters in the Georges River, Sydney, NSW. It also shows that genetic diversity is lower in wild populations as compared to farmed populations. Mean observed heterozygosity in wild populations was found to be 0.213, whereas in farmed oysters it was 0.243. The level of inbreeding (0.25) in wild populations was found to be significant. Two distinct genetic clusters of oysters were revealed by a Discriminant Analysis of Principal Components. This lack of gene flow between the wild and farmed oysters and the high level of inbreeding may be due to severe disease outbreaks in wild populations where there is little resilience to both QX and winter mortality disease, or the overharvesting of wild populations during the establishment of breeding programs. In this case, the introgression of the genetic material of the farmed Sydney Rock Oyster back into its wild counterparts could have the beneficial effect of increasing genetic diversity, as well as increasing resilience to disease and other environmental stressors.

# Introduction

Oysters are an environmentally significant estuarine mollusc species who provide several ecosystem services such as the provision of reef structure, and reducing water turbidity [1]. Their presence is also used to determine water quality [2]. They are of great economic importance, with oyster farming being the largest aquaculture industry in the state of NSW [3, 4]. Sydney Rock Oysters (SROs) are the most intensively farmed oyster in Australia [5]. Their economic significance has led to intense efforts to develop oysters that are resistant to both Queensland unknown (QX) and winter mortality disease. Both diseases have, in the past, resulted in greater than 80% declines in farmed SROs in certain areas.

The Sydney Rock Oyster is classed as functionally extinct in the wild [6] and therefore no longer capable of providing ecosystem services [1], however, populations do persist in the Georges River, at Wallis Lake and at Port Stephens [7]. The Georges River has been severely affected by outbreaks of QX disease [4], therefore, programs that selectively breed SROs here have developed disease resistant stock. Broodstock here consist of survivors of QX outbreaks from each successive generation [4]. As a result, SROs in the Georges River demonstrate higher levels of disease resilience as well as faster growth and increased resilience to environmental stressors [6].

Historically, oyster spat (juveniles) for use in breeding programs have not always been sourced from local estuaries. This has likely impacted the genetic relationship between wild and farmed oysters. The intensive selective breeding that has occurred within the oyster industry is also likely to have resulted in differences between wild and hatchery farmed SROs. However, both wild populations and farmed populations of SROs inhabit the Georges River.  
  
This report hypothesises that genetic introgression is occurring between hatchery reared and wild oysters in the Georges River, NSW, due to their proximity to one another and their broadcast spawning method of reproduction. It also hypothesises that there will be greater genetic variability in wild populations than in hatchery reared populations of the SRO. This is thought to be the case due to the intense breeding regimes farmed SROs are subject to, selecting only for alleles that confer several beneficial traits.

# Methods

The data used in this study were obtained from “Thompson, J.A. et al. (2017) Lack of genetic introgression between wild and selectively bred Sydney rock oysters Saccostrea glomerata”, and were provided to this author personally by J.A Thompson.

Sampling and collection of oysters Wild and farmed oysters were sourced from two sites on the Georges River: Woolooware Bay (34°02’14.2’’ S 151°08’51.5’’ E) and Quibray Bay (34°01’29.7’’ S 151°10’50.3’’ E), NSW. At each site, three groups of oysters were sampled, (n=12 per group). The groups were wild catch (WC), collected approximately 1km from any hatchery and from an area of shoreline spanning 10 – 20 metres; overcatch (OC), sampled from pylons supporting oyster farms; and B2, the line of selectively bred, 6th generation oysters (supplied by NSW DPI). At both sites, oysters suffer mortality from both winter mortality disease and QX disease. Quibray Bay is affected more severely by winter mortality disease, and Woolooware Bay more affected by QX disease. Two reference groups were sourced from Port Stephens Hatchery, Port Stephens, NSW (32° 44’ 12.5’’ S 152° 3’ 18.9’’ E). These were 6th generation hatchery reared B2 oysters (HB2, n = 11) and hatchery control (HC, n = 10) oysters, which have not been selectively bred. Oysters (n=93) were shucked, then had their gill tissue cut out and stored in 95% ethanol to prepare it for processing.

## Sequencing and genotyping

All sequencing was performed by Diversity Arrays Technology (DArT). The DArTSeq™ protocol was used to identify single nucleotide polymorphisms (SNPs). DNA was extracted and then purified, then amplified by PCR primers specific to the barcode sequences. The final dataset was filtered for missing data and only oysters with greater than 95% data were used. The data was then converted into .GenAlEx format [6]. Outliers (n=3) were removed by J.A Thompson for the purposes of this study.

## Analysis

The data consist of 1200 SNPs for 90 individuals. Data were provided to this author in .GenAlEx and .GEN (Genepop) format as two separate files: one partitioned by cluster (Wild/B2, n=2) and one partitioned by population (n=8).

The dependent variables in the study are the allele frequencies in the form of observed heterozygosity (Ho), expected heterozygosity (Hs), fixation index (FST) and inbreeding coefficient (F/FIS). Independent variables are cluster (levels: wild/selectively bred), population (8 levels), and locus (1200 levels).

The null hypotheses state that 1) introgression between wild and selectively bred (B2) oysters is not occurring in the Georges River and that 2) there will be lower genetic variability in wild populations than in B2 populations of oysters. To test these hypotheses, all statistical analyses were performed using the software package R version 4.0.5 (R Development Core Team 2015). Data were transformed from a Genepop into a genind object using the ‘read.genepop’ function in the adegenet package 2.1.3 [8]. All statistics were computed on 90 individuals at 1200 loci. Summary statistics on clusters and populations were calculated using the ‘basic.stats’ function in hierfstat 0.04-22 [9]. A one-sided t-test was performed to test for inbreeding. The t-test assumes normality of variance, tested for by a qqnorm graph.

To determine the extent to which genetic material was shared between the eight populations, global Fst and pairwise Fst values with accompanying 95% CIs were calculated using 100 bootstraps with the function ‘diffCalc’ in the diveRsity package 1.9.90 [10]. Variation within the populations was analysed using a Principal Component Analysis (PCA). Assumptions of PCA include multiple variables at continuous levels, a linear relationship between all variables, no significant outliers, sampling adequacy, and adequate correlations suitable for data reduction. A linear relationship was tested for using the ‘qqnorm’ function in the stats package 4.0.5 [11]. A PCA scatter of the eight population IDs with 95% confidence intervals was constructed using the ‘dudi.pca’ function in the poppr package 2.9.0 [12]. A PCA was also performed to detect variation between the clusters (wild/B2). This PCA plot was constructed using the function ‘indpca’ in the adegenet package 2.1.3. A Discriminant Analysis of Principal Components (DAPC) was also performed to test for variance between the populations and determine the number of genetically distinct clusters. This was performed on the population IDs, not the clusters, so the number of clusters was not known. The function ‘find.clusters’ was used to determine the number of genetically distinct clusters (K). The DAPC was then calculated with K, using the function ‘dapc’ in adegenet 2.1.3. Finally, a neighbour joining tree was constructed to determine the genetic distance between populations. The neighbour joining tree was calculated using the function ‘nj’ in the ape package 5.4-1 [13].

# Results

Both the null hypotheses are retained. This study has shown that introgression is not occurring between wild and selectively bred oysters in the Georges River. It has also shown that there is lower genetic variability in wild populations than in B2 populations of oysters. Summary statistics (Table 1) show that HO across all populations ranged from 0.193 to 0.263, HS from 0.268 to 0.285 and FIS from 0.07 to 0.253. Observed heterozygosity for all of the B2 populations was higher (0.242) than for the pooled wild populations (0.213) (Table 2). The inbreeding coefficient was higher in the pooled wild population (FIS=0.250) than in the B2 line (FIS=0.155). The t-test for inbreeding coefficient F showed a significant deviation from a mean of zero (F = 0.249, t = 37.449, df = 1199, p < 2.2e-16, 95% CI: 0.236 - 0.262). The assumption of normality for the t-test was met, visualised by the normal distribution seen in the QQ norm plot (Figure 1).

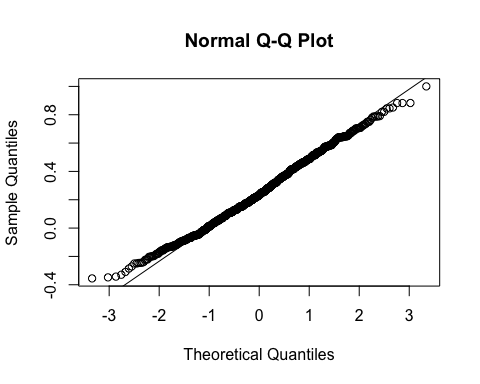
**Table 1.** Summary statistics giving mean HO, HS and FIS for each population. Inbreeding is lower in all B2 populations, and observed heterozygosity is lower in all wild population.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **N** | **HO** | **HS** | **FIS** |
| **QB2** | 9 | 0.263 | 0.285 | 0.070 |
| **WB2** | 12 | 0.235 | 0.273 | 0.127 |
| **HB2** | 11 | 0.231 | 0.282 | 0.167 |
|  |  |  |  |  |
| **QOC** | 12 | 0.203 | 0.278 | 0.237 |
| **QWC** | 12 | 0.193 | 0.268 | 0.253 |
| **WOC** | 12 | 0.241 | 0.278 | 0.123 |
| **WWC** | 12 | 0.219 | 0.275 | 0.195 |
| **HC** | 10 | 0.203 | 0.277 | 0.232 |

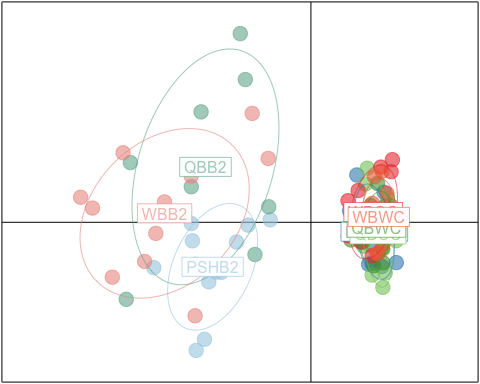
**Table 2.** Summary statistics of the mean for all HO, HS and FIS across the two clusters.

|  |  |  |  |
| --- | --- | --- | --- |
|  | **HO** | **HS** | **FIS** |
| **B2** | 0.242 | 0.283 | 0.155 |
| **Wild** | 0.213 | 0.277 | 0.250 |

##   
## One Sample t-test  
##   
## data: (div$Hexp - div$Hobs)/div$Hexp  
## t = 37.449, df = 1199, p-value < 2.2e-16  
## alternative hypothesis: true mean is not equal to 0  
## 95 percent confidence interval:  
## 0.2360068 0.2621024  
## sample estimates:  
## mean of x   
## 0.2490546

   
**Figure 1.** QQnorm graph showing the normal distribution of data points.

The PCA performed on the populations (Figure 2) shows a distinct genetic differentiation between the wild and B2 populations. It also shows greater variability within the B2 populations than within wild populations. Quibray Bay B2 oysters have the highest level of genetic diversity, shown by the size of the oval which represents the 95% CI, and the HO value of 0.263. The PCA performed on the cluster data (Figure 3) shows the same variance, with separation between the two clusters and high variability within the B2 cluster itself. The DAPC produced two genetically distinct clusters with all B2 oysters pooled into Cluster 1 (n = 32) and all wild pooled into Cluster 2 (n = 58). The DAPC showed no overlap between the two clusters (K=2, Bayesian Information Criterion =437.64) (Figure 4).

Chart, scatter chart

Description automatically generated  
  
   
FST values were 0.05 at the population level (95% CI: 0.036-0.067) and 0.233 at the individual level (95% CI: 0.2035-0.2570). Pairwise FST values were calculated at 100 bootstraps.

**Figure 3.** PCA performed on cluster data, showing distinction between the B2 and wild clusters.

**Figure 2.** PCA performed on population data, showing two distinct clusters (wild and B2), as well as substantial differentiation within the B2 cluster itself. PCA on the first two principal components.

## pop Ind  
## Total 0.0500853 0.2329216  
## pop 0.0000000 0.1924765

## [1] 0.0500853

Pairwise FST results revealed that all wild-wild and all B2-B2 comparisons have lower pairwise FST values with all 95% CIs containing zero, while all wild−B2 sample group pairs have higher pairwise FST values and CIs not containing zero (Table 3). The neighbour joining tree was constructed using the pairwise FST values and again shows two genetically distinct clusters with no overlap between populations (Figure 5).  
  
**Table 3.** FST values for population. Bold type indicates 95% confidence intervals that do not contain zero and therefore show that gene flow is not occurring.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | HB2 | HC | QB2 | QOC | QWC | WB2 | WOC | WWC |
| HB2 |  |  |  |  |  |  |  |  |
| HC | **0.0694** |  |  |  |  |  |  |  |
| QB2 | 0.0227 | **0.073** |  |  |  |  |  |  |
| QOC | **0.0714** | 0.0037 | **0.0745** |  |  |  |  |  |
| QWC | **0.0723** | 0.0039 | **0.0735** | 0.0034 |  |  |  |  |
| WB2 | 0.0238 | **0.0997** | 0.0152 | **0.1041** | **0.1054** |  |  |  |
| WOC | **0.0694** | 0.0097 | **0.0720** | 0.0080 | 0.0115 | **0.1081** |  |  |
| WWC | **0.0753** | 0.007 | **0.0731** | 0.0033 | 0.0036 | **0.1035** | 0.0088 |  |

Chart, histogram

Description automatically generated  
  
**Figure 4.**  DAPC showing two distinct genetic clusters. Cluster 1 shows   
the B2 population (n=32), Cluster 2 shows the wild population (n = 58).   
There is no genetic overlap.   
  
A picture containing chart

Description automatically generated

**Figure 5.** Neighbor joining tree based on pairwise FST calculations.

# Discussion

The hypothesis that introgression is occurring between wild and selectively bred oysters in the Georges River is rejected. Both PCAs show a distinct level of genetic differentiation between the B2 and wild clusters. The DAPC independently identified two distinct clusters in the pool of eight populations. It can therefore be inferred that introgressive gene flow has not occurred between any population of B2 and wild oyster. The t-test shows that inbreeding is significant in the wild cluster, further showing that limited gene flow has occurred and that clusters are separate. Low FST values and CIs that do not contain zero in the pools of wild and B2 oysters show that there are significant levels of gene flow within the clusters, but not between. The neighbour joining tree shows the same.

The hypothesis that there is greater genetic variability in wild oyster populations than in the B2 line must also be rejected. The PCA performed on the eight populations shows higher levels of genetic variability within the B2 lines, with the highest level of differentiation occurring in the Quibray Bay B2 population. Observed heterozygosity was higher in the B2 cluster than in the wild cluster. The higher degree of inbreeding in the wild cluster aligns with the lower level of observed heterozygosity and once again demonstrates a lack of gene flow between wild and B2 oysters across all locations.

The results obtained in this study are very closely aligned with those in the original study by J.A. Thompson [6], and the inferences drawn are similar, too.

The lack of introgression occurring between wild and B2 populations may be as a result of factors concerning reproductive success. It is possible that B2 spat is less viable than wild spat, and therefore, does not survive for a long in the water column. Also, the ability for cross-fertilisation between wild and B2 oysters may be limited after generations of selective breeding; potentially B2 spat is no longer compatible with wild spat. It is also possible that the distances between wild and B2 oysters sampled in this study (~1km) are too great for fertilisation to occur on a scale large enough to detect here. Lastly, it has been noted that gravid oysters are often harvested before they have a chance to broadcast spawn as they are largest during this period. This therefore removes a substantial amount of genetic material from the fertilisation events [6]. All of these factors would contribute to the lack of gene flow between the two clusters.

The low levels of genetic diversity within the wild oyster populations have been attributed to several factors. Both Woolooware Bay and Quibray Bay are subject to outbreaks of QX and winter mortality disease [4], respectively, resulting in possible population bottlenecks, thus reducing the genetic variation within the populations. B2 oysters are specifically bred with disease resilience and are therefore not subject to the same pressures or bottlenecks. Wild oysters have also potentially been historically overharvested [6] for use in breeding programs, therefore suffering from further loss to genetic diversity within their populations. Large, disease resilient oysters have been harvested from wild populations meaning that genetically robust individuals are no longer able to contribute to the diversity within the gene pool. This has left wild populations functionally extinct, with high levels of inbreeding and low levels of heterozygosity.

Small sample sizes cannot be ruled out as the possible cause of the rejection of both hypotheses in this study. It is possible that introgression is occurring on small scales, but that this study failed to detect it [6]. Future studies could ensure they sample greater numbers of oysters.

In the case of the oyster populations in the Georges River, introgressive gene flow back into wild populations has the potential to restore (over substantial periods of time) the functionality of the oyster reefs by increasing size, resilience to disease and resistance to detrimental effects of environmental changes.

# References

1 Coen, L.D. et al. Ecosystem services related to oyster restoration. , Marine Ecology Progress Series, 341. (2007)   
2 Melwani, A.R. et al. (2016) Differential proteomic response of Sydney rock oysters (Saccostrea glomerata) to prolonged environmental stress. Aquat. Toxicol. 173, 53–62   
3 Rankin, C. et al. (2018) Shell shape and meat condition in selectively bred Sydney rock oysters, Saccostrea glomerata (Gould, 1850): The influence of grow-out methods. Aquac. Res. 49,   
4 Vaibhav, V. et al. (2016) Biomarkers of Winter Mortality resistance in selectively bred Sydney rock oysters (Saccostrea glomerata). Aquaculture 465,   
5 Nguyen, V.K. et al. (2021) Dynamics of the Sydney rock oyster microbiota before and during a QX disease event. Aquaculture 541,   
6 Thompson, J.A. et al. (2017) Lack of genetic introgression between wild and selectively bred Sydney rock oysters Saccostrea glomerata. Mar. Ecol. Prog. Ser. 570,   
7 Parker, L.M. et al. (2011) Populations of the Sydney rock oyster, Saccostrea glomerata, vary in response to ocean acidification. Mar. Biol. 158,   
8 Jombart, T. (2008) Adegenet: A R package for the multivariate analysis of genetic markers. Bioinformatics 24,   
9 Goudet, J. (2005) HIERFSTAT, a package for R to compute and test hierarchical F-statistics. Mol. Ecol. Notes 5, 184–186   
10 Keenan, K. et al. (2013) DiveRsity: An R package for the estimation and exploration of population genetics parameters and their associated errors. Methods Ecol. Evol. 4, 782–788 11 R Core Team R: A language and environment for statistical computing. , R Foundation for Statistical Computing. (2019)   
12 Kamvar, Z.N. et al. (2014) Poppr: An R package for genetic analysis of populations with clonal, partially clonal, and/or sexual reproduction. PeerJ 2014,   
13 Paradis, E. et al. (2004) APE: Analyses of phylogenetics and evolution in R language. Bioinformatics 20.