Population structure of Ballot's saucer scallop (*Ylistrum balloti*) for the east coast of Queensland.

Technical Report

September 2024

Queensland Government

This publication has been compiled by G. Scata, J.G. McGilvray, K.D. Masci, S.M. Thomas of Fisheries Queensland, Department of Agriculture and Fisheries.

Enquiries and feedback regarding this document can be made as follows:
Email: info@daf.qld.gov.au
Telephone: 13 25 23 (Queensland callers only)
(07) 3404 6999 (outside Queensland)
Monday, Tuesday, Wednesday and Friday: 8 am to 5 pm, Thursday: 9 am to 5 pm
Post: Department of Agriculture and Fisheries GPO Box 46 BRISBANE QLD 4001 AUSTRALIA
Website: daf.qld.gov.au

Interpreter statement
The Queensland Government is committed to providing accessible services to Queenslanders from all culturally and linguistically diverse backgrounds. If you need an interpreter to help you understand this document, call 13 25 23 or visit daf.qld.gov.au and search for 'interpreter'.

Contents

Executive Summary

Ballot's saucer scallops *(Ylistrum balloti)* are harvested by the east coast otter trawl fishery and contribute a higher proportion of the overall yearly catch than the co-located mud scallop *(Amusium pleuronectes).* Recent estimates of Ballot's saucer scallop biomass are below the defined limit reference point (20%), and the fishery is currently closed across much of its area (limited fishing is allowed in the southern offshore management region). An underlying assumption for fishery management is that Ballot's saucer scallops along Queensland's east coast are one population to be managed and assessed as a single biological stock. Significant aggregations of Ballot's saucer scallops have been detected in the central trawl management region despite the low biomass estimates. Trawl fishery representatives have concerns about the single stock approach. Their observations of morphometric and colour differences between Ballot's saucer scallops from the central trawl management region and the management regions to the south are suggestive of distinct populations. To determine the population structure Fisheries Queensland used High Throughput Sequencing to identify 3217-5754 Single Nucleotide Polymorphisms (SNPs). The results indicated small (F_{STs} ~ 0.005), but significant ($p<0.001$) differences between scallops either side of 22 degrees South and was further supported by PCA, AMOVA and Admixture analyses. The study shows the existence of two populations of Ballot's saucer scallops on the east coast of Queensland.

Acknowledgements

The authors would like to thank Sue Helmke, Malcolm Pearce, Sian Breen and Thomas Hatley, from Fisheries Queensland for supporting and funding this study.

To the commercial fishers who collected the samples and to the Fishery Monitoring team who completed the laboratory processing of those samples and contributed to the final report we also extend our thanks.

For their guidance and advice during analysis, including detailed interpretations of results and well-informed comments on the draft report the authors extend their thanks to Jessica Morgan and Sam Williams of Agri-Science Queensland.

Introduction

Two species of scallop, Ballot's saucer scallop *(Ylistrum balloti)* and mud scallop *(Amusium pleuronectes)* are harvested on the east coast of Queensland. Ballot's saucer scallop has the highest harvest of the two species which is managed under approved trawl fishery regional harvest strategies. Ballot's saucer scallops are a secondary species in the central, and southern offshore trawl regions, (DAF[a] 2021, DAF[b] 2021). In the southern inshore trawl region, they are a primary target species (DAF[c] 2021). These three management regions include fishable waters from Townsville south to the Queensland – New South Wales border.

The most recent stock assessment (using data up to the end of 2022) estimated Ballot's saucer scallop biomass to be 15% of pre-1956 levels (French 2023), with no significant increase in the biomass estimate from 12 months previously (Wortman 2022). Based on these assessments, and to rebuild the stock, harvest of Ballot's saucer scallops is not permitted in central and southern inshore management regions, although a limited opening of the fishery is permitted in the southern offshore trawl region.

Ballot's saucer scallops spawn in the austral winter and have a life history that includes a significant larval phase before settlement followed by fast growth to maturity (Dredge 1981). The long larval phase in dynamic oceanic waters would suggest there is a high probability that scallops form a single population along Queensland's east coast. Miller et al (2013) used 11 microsatellites to determine population structure of pipi *(Dontax deltoides)* and concluded that ocean currents, an extended spawning season and long larval phase were the key factors driving connectivity along a 1500 km stretch of Australia's east coast. Schilling et al (2022) concluded connectivity of spanner crabs *(Ranina ranina)* between Southern Queensland and New South Wales was a result of larval dispersal driven by the East Australia Current. A single population of spanner crabs *(Ranina ranina)* had been previously confirmed by Brown et al (1999), using mitochondrial DNA.

O'Brien et al (2005), used seven microsatellite loci to conclude that Ballot's saucer scallops comprised a single population on the east coast of Queensland. Courtney et al (2015) used larval dispersal modelling to demonstrate connectivity between Ballot's saucer scallops from Yeppoon to Double Island Point, however areas to the north of Yeppoon were not considered in the study.

More recently, McMillan et al (2024), used 3,031 Single Nucleotide Polymorphisms (SNPs) (loci) to demonstrate that the co-located species, Sand Bug (*Thenus australiensis),* also formed a single population on the east coast of Queensland.

Understanding temporal and spatial population structure is key for successful management and assessment of marine resources. Individual biological stocks are formed when interbreeding individuals become reproductively isolated from other groups. Begg et al (1999), summarises the methodologies that can be used to identify population structure, including catch data, tagging studies, meristics, morphometrics, scale morphology, parasites, cryogenics, protein electrophoresis, otolith elemental composition, stable isotope measurements, otolith microstructure, shape analysis and thermal marking. Mitochondrial DNA and nuclear DNA are also integrated into population structure investigations and methodologies in this area are continually evolving.

During 2023, trawl fishers reported significant aggregations of Ballot's saucer scallops in the Townsville and Hydrographers Passage areas, which fall within the central management region. These aggregations are approximately 90 nm and 250 nm, respectively, to the north of the northern extent of the traditional scallop fishing grounds. Aggregations in these areas are not uncommon and in the mid 1990's they collectively yielded single year harvests of more than 300T. Despite these productive years, the harvest from these patches rarely contributes more than 10% of the overall east coast scallop fishery harvest (French 2023). An average harvest of ~ 30T from these areas between 2000 and 2020 does represent an economic benefit for the local trawl operators who believe that the scallops in these aggregations are a different stock due to morphometric (smaller overall shell size) and colour (lighter) differences and argue that scallops in these aggregations should be managed under a different regime with a limited harvest.

Fisheries Queensland aims to address industry concerns by using advances in genome sequencing technology and analysis to re investigate the population structure of Ballot's saucer scallop on Queensland's east coast.

Methods

Sampling

Samples of 30 individuals, greater than 90mm shell height were targeted from locations in Table 1, which are known areas of scallop. These areas were also sampled by O'Brien et al (2005). Four commercial fishers were authorised under general fisheries permit (GFP 213514) to collect samples from the selected areas between November 2023 and February 2024 (Figure 1).

Table 1 – Summary of industry-based Ballot's saucer scallop sample collection. #Fraser is officially known as K'gari.

Management region	Location	Cfish grid	Date	n individuals
Central Area	Lucinda Region	K20	10/11/23	30
Central Area	Mackay Region	R25	29/10/23	30
Southern Inshore	Rockhampton Offshore	S29	2/11/23	29
Southern Inshore	Rockhampton Offshore	T30	10/11/23	30
Southern Inshore	Hervey Bay	V32	13/2/24	30
Southern Offshore	Fraser# Offshore	W34	4/12/23	30

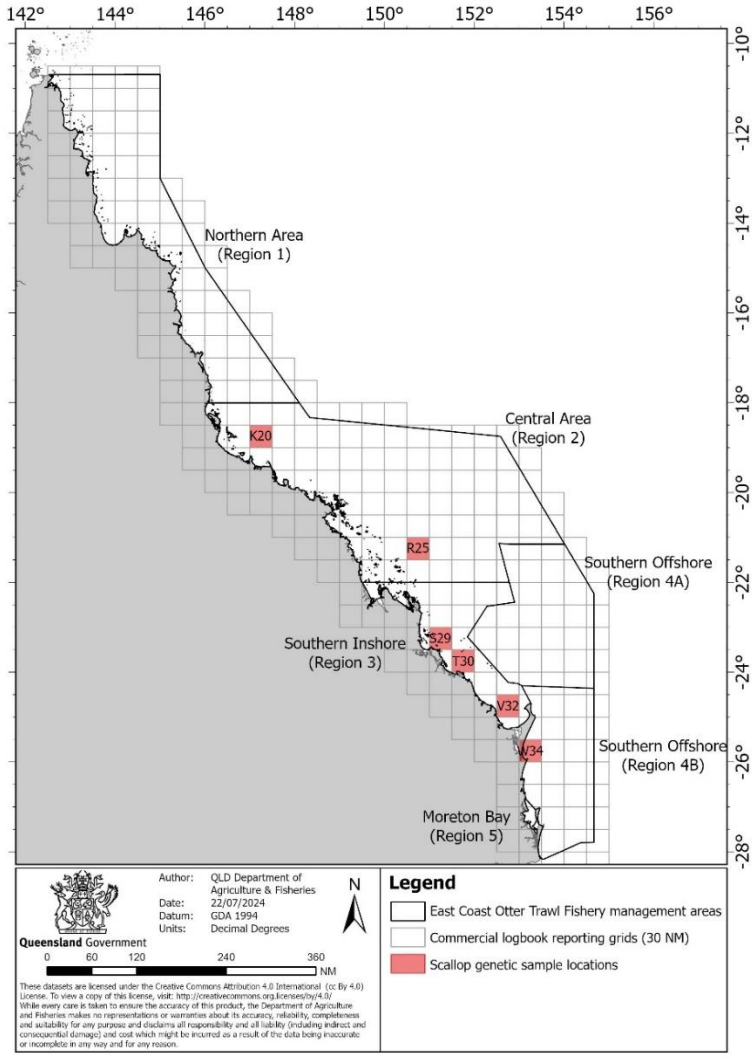

Figure 1. Sample collection locations with 30nm CFish grids and East Coat Otter Trawl fishery management regions.

Samples were processed between December and March 2024, Appendix 1. Samples were thawed at ambient temperature in a single layer, using a basket which allowed fluid to drain and avoid potential sources of cross contamination. The following data was recorded from each individual scallop –

- Catch location
- Date caught
- Shell height
- Total weight
- Meat weight
- Gonad weight
- Macroscopic gonad stage

A piece of tissue was removed from the adductor muscle and stored in 100 per cent molecular grade ethanol with a unique sample number. Samples were then stored at –20 degrees Celsius.

Tissue samples were further subsampled to 10-15mg, placed in 100% molecular grade ethanol in 96-well plates in preparation to send to Diversity Arrays Technology (DArT). Samples were added to the plate so that an individual from each region was present in the first 2 columns of the first plate and sample from nearby locations were separated as much as possible to limit adverse batch effects during the development of the reduced genome library.

Tissues samples from 179 individuals from 6 different sampling locations (Table 1) were sent to DArT for DNA extraction, DNA digestion using two restriction enzymes (PstI and HpaII), adapter ligation (ligation of Illumina primers with individual barcodes) and amplification of fragments, followed by High-Throughput Sequencing.

The resulting sequences were processed by DArT using proprietary analytical pipelines to filter away low-quality sequences, assemble loci and call genotypes for the 179 individuals. Individuals are therefore characterised by a set of loci (DNA sequences at specific locations within the genome) containing one or more SNPs (genetic markers with two variants or “alleles”) for which they are coded as 0 if homozygous for the reference allele (major allele), 1 if heterozygous and 2 if homozygous for the SNP allele (minor allele).

Filtering

Quality filtering of the SNPs data and population structure analyses were performed in R Statistical Software (v 4.2.1; R Core Team, 2022) using mainly packages -

- dartR 2.7.2 and dartRverse (Gruber et al., 2018)
- outflank (0.2) (Whitlock & Lotterhos, 2015)
- PCAdapt (4.3.5) (Luu et al., 2017)
- SNPRelate (1.36.1) (Zheng & Zheng, 2013)
- Radiator (1.3.3) (Gosselin, 2020)
- poppr (Kamvar et al., 2014)
- LEA (3.8.0) (Frichot & François, 2015)

Following a Principal Component Analysis (PCA) (Figure 1, Appendix 2), quality filtering to remove missing data (low quality SNPs or individuals) or SNPs representing non-target DNA (i.e. mitochondrial, repeated sequences) was undertaken. The dataset was also analysed for the presence of sex-linked loci (Robledo-Ruiz et al., 2023).

Six filtering parameters at different thresholds were then tested on the original dataset of 40,162 loci and 179 individuals. These included in order:

1. Reproducibility
2. Read Depth
3. Individual Call Rate (individual missingness)
4. Monomorphic loci
5. Locus Call Rate (Locus missingness)
6. Secondaries SNPs

Table 2 provides an explanation of each quality filtering parameter and the thresholds tested and applied for each.

Table 2 - Filtering parameters applied with thresholds tested for each parameter

Filtering parameter	Threshold tested	Threshold applied to final datasets	Purpose
Reproducibility (Rep)	0.98, 0.95	0.98, 0.95	Identify reproducibility across 30 replicated samples. Removes poorly sequenced or poor-quality loci.
Average Read Depth (rDepth)	5-125	5-125	Checks average number of sequence-tag copies of a locus. Removes false homozygotes (due to low sample size or nonamplification of an allele), sequencing errors, multi-locus contigs, paralogue loci, mtDNA and rDNA, repetitive sequences.
Individual Call Rate (Ind CR)	0.68	0.68	Identifies % of missing data per individual. Removes poorly sequenced/ poor-quality individuals (or individuals with very different loci).
Monomorphic	-	Removed all	Removes monomorphic loci (uninformative), that can be created when removing individuals or populations from the dataset.
Locus Call Rate (Loc CR)	0.95, 0.80	0.95, 0.80	Identifies % of missing data per locus. Removes poorly sequenced/ poor-quality loci.
Minor Allele Frequency (MAF)	0.04	0.04	Removes PCR errors and/or rare alleles.
Secondary SNPs (2nd)	-	Keep the SNP with best reproducibility and “Polymorphism information content” (informativeness)	Checks for sequenced fragments with more than one SNP within the same locus. Removes physically linked loci (less informative/can provide false signals of population structure).
Hardy Weinberg Equilibrium (HWE)		Removed loci out of HWE across all sampling groups analysed independently (used FDR-BH approach for multiple tests correction of p-values)	Removes loci out of HWE, which are likely to be under selection and do not satisfy the assumptions of the downstream population structure analyses.
Heterozygosity (Het)	-	None applied	Removes potential mixed samples (individuals with extremely high heterozygosity).
Outliers loci (Outl)		Removed loci detected as outliers concurrently by two methods, Outflank and PCAdapt	Removes loci large Fsts compared to other loci (likely under selection). These loci could bias downstream population structure analyses.

The threshold combinations in Table 2, resulted in four preliminary datasets (Table 4, datasets 1–4). A PCA was conducted on these preliminary datasets to assess the effects of the filtering steps. As a

similar clustering pattern was observed across all datasets, the most stringent and the most relaxed (datasets 1 and 4), were selected for further filtering.

These datasets (datasets 1 and 4) were assessed for deviation from Hardy-Weinberg Equilibrium, for individual and population heterozygosity, population F_{is} (inbreeding coefficient), genetic relatedness and for outlier loci (potentially under selection).

Deviations from Hardy-Weinberg Equilibrium were assessed using Fisher's Exact test (α=0.05) with p-values corrected for multiple testing through the False Discovery Rate (FDR) approach applying the method of Benjamini and Hochner (FDR-BH) (package dartR::gl.report.hwe).

Individual and population heterozygosity and F_{is} were investigated through dartR (dartR ::gl.report.heterozygosity).

Individuals pairwise genetic relatedness was assessed by estimating the coancestry coefficient through both the Maximum Likelihood and King-Robust methods implemented in the SNPRelate R package (Milligan, 2003; Zheng et al., 2012; Zheng and Zheng, 2013) (SNPRelate::snpgdsIBDMLE or ::snpgdsIBDKING were used with default parameters except for "kinship=TRUE" in snpgdsIBDMLE and "autosome.only=FALSE").

To detect duplicate genomes pairwise genetic distances (Manhattan distance) between individuals were computed with the package Radiator (Radiator::detect_duplicate_genomes, used with default parameters).

Detection of F_{ST} (genetic fixation index) outlier loci was conducted with the programs OUTFLANK (Whitlock and Lotterhos, 2015) and PCADAPT (Luu et al, 2017). In OUTFLANK, multiple tests corrections were carried out using the False Discovery Rate (FDR) correction of Storey and Tibshirani (2003). Parameter settings in OUTFLANK were -

- FDR q-value of 0.05;
- minimum heterozygosity 0.1;
- trimming 5% loci from both tails of the distribution.

In PCADAPT, the optimal number of Principal Components were identified as those with the potential to explain population structure (taken from the bend in the eigen values plot). First, PCADAPT was conducted retaining 50 PCs and the resulting screeplot plot (PCs plotted against the percentage of their explained variance) was used to determine the optimal number of PCs to retain in the final PCADAPT. The optimal number of PCs to retain corresponds to the largest value of K before the plateau in 'scree plot' (elbow method). The PCs at the left of the elbow in the screeplot represent PCs explaining the majority of variance, the remaining PCs represent mostly inter-individual variance or noise. In addition, the optimal number of PCs to retain was also confirmed by plotting the PCADAPT results with sequential PCs (PC1 vs PC2, PC2 vs PC3) until no further structure was obvious in the plot. The optimal number of PCs to retain is the largest number of PCs which still reveal population structure. Significant outliers were detected applying multiple tests corrections with the FDR-BH method. Only loci identified as potential outliers by both methods were removed.

These final filtering steps led to the final datasets 5 and 6 (Table 4).

Population structure analyses

Analyses for population structure were conducted on either one or both of the final datasets. These consisted of a final PCA (both datasets), pairwise population F_{STs} (both datasets), genetic Admixture analyses (both datasets), spatial autocorrelation, AMOVA and k-means clustering (only on the most stringent dataset) (R package *adegenet*, Jombart et al., 2018).

Principal Component Analysis

A Pearson Principal Component Analysis (PCA) was conducted through the package dartR (dartR::gl.pcoa).

Pairwise FSTs among sampling groups

Pairwise F_{STs} were estimated across all sampling groups to test for significance and magnitude of genetic differentiation among sampling groups. Pairwise F_{STs} were calculated with dartR (gl.fst.pop). Significance values (the probability of F_{ST} values to be different from zero) were obtained through 1000 bootstraps and 95% confidence intervals were calculated with this method. To account for multiple testing, only p-values <0.001 were considered significant.

Admixture analysis

Analysis of genetic admixture was performed through Sparse Nonnegative Matrix Factorization (SNMF) algorithms implemented in the LEA R package (Frichot & François, 2015). The SNMF algorithm was run for K values ranging from 1 to 8 (higher than the number of sampling groups) with 10 repetitions per each K, and all other parameters set as the default value. The most likely number of populations (K stocks) was selected as the number of K showing the lowest entropy criterion (Frichot et al., 2014).

Analysis of Molecular Variance (AMOVA or hierarchical F_{STs})

The AMOVA was conducted with the package poppr through the implementation of the package pegas, (Kamvar et al., 2014) using either Nei's genetic distance or simple Euclidean distance as pairwise distance between individuals, with 100000 permutations and with "farthest neighbour" clustering algorithm (default). Nei's genetic distance accounts for the impact of the emergence of mutations along with genetic drift (random fluctuations in alleles frequencies in a population), on population differentiation among populations, and, assuming balance between drift and mutation, increases proportionally with time after divergence of the two populations. Populations with longer periods of separation will have larger values of Nei's genetic distances (Georges et al., 2023).

The AMOVA analysis was conducted by defining three levels of clustering: i) individuals, ii) populations, iii) the two clusters identified with both PCA and pairwise F_{ST}s.
The AMOVA was conducted only for the final dataset, the most stringent filtering scenario (Table 4, dataset 5).

Results

Filtering

The original dataset obtained from DArT contained 40,162 SNPs over 179 individuals.

No sex-linked loci were detected. Five individuals with low call rate (missing over 30% of the loci) were removed (individual call rate threshold 0.68). The other filters applied are reported in Table 3 with the number of individuals and SNPs removed at each step.
Application of the filtering steps outlined in Table 3, lead to the creation of four preliminary datasets (Table 4, dataset 1-4).

The PCA conducted on these preliminary datasets showed similar results to the PCA on the pre-filtered dataset, with two clusters identified:

i) Mackay and Lucinda,
ii) Rockhampton Offshore, Hervey Bay and Fraser Offshore

These clusters were separated along the primary PCA axis which explained about 1.4% of the variation. Two of these preliminary scenarios, the most stringent and the most relaxed (Table 4, datasets 1 and 4), were selected to continue with downstream analyses.

For the most stringent dataset (Table 4, dataset 1), no loci were found out of HWE in all populations, while for the most relaxed dataset (Table 4, dataset 4) only one locus was found out of HWE in all populations, but nine loci were out of HWE in five out of six populations, so these ten loci were removed (as they were deviating from HWE most populations).

No individuals with outlier heterozygosity (i.e. potential contamination) were detected and no related or duplicate individuals were found in the two datasets (Table 4, dataset 1 and 4).

Only loci identified as outliers by both Outflank and PCAdapt were removed. For PCAdapt, in both datasets 1 and 4, 2 PCs were retained, as additional PCs did not reveal any further population structure. In the most stringent dataset (Table 4, dataset 1), 12 outlier loci were removed while in the most relaxed dataset (Table 4, dataset 4), 15 outlier loci were removed. These final filtering steps led to two final datasets: stringent and relaxed (Table 4, datasets 5 and 6).

Table 3. Filtering thresholds applied for each parameter and number of Single Nucleotide Polymorphisms (SNPs) and individual removed for each filtering step.

Filter name	Filter applied	n SNPs retained	n SNPs removed	n Ind retained	n Ind Removed
Raw data		40162	-	179	0
Monomorphic		40162		179	0
Sex.linked (xy)	none	-	-	-	-
Reproducibility	**Rep 0.98** **Rep0.95**	35211 38502	4951 1660	179 179	0 0
Read Depth	**rDepth 5-125** • Rep 0.98 • Rep 0.95	 29704 32302	 5507 6200	 179 179	 0 0
Individual CR (remove new monomorphic loci created)	**IndCR >0.68** • Rep 0.98 • Rep 0.95	 29456 32054	 248 248	 174 174	 5 5
Locus CR	**LocCR0.80** • Rep 0.98 • Rep 0.95 **Loc CR0.95** • Rep 0.98 • Rep 0.95	 18839 20525 12370 13354	 10617 11529 17086 18700	 174 174 174 174	 0 0 0 0
MAF	**MAF0.04** • Rep 0.98-LocCR0.80 • Rep 0.98-LocCR0.95 • Rep 0.95-LocCR0.80 • Rep 0.98-LocCR0.95	 5493 3459 6987 4321	 13346 8911 13538 9033	 174 174 174 174	 0 0 0 0
Secondaries	**Secondaries** • Rep 0.98-LocCR0.80 • Rep 0.98-LocCR0.95 • Rep 0.95-LocCR0.80 • Rep 0.95-LocCR0.95	 4584 3139 5779 3915	 909 320 1208 406	 174 174 174 174	 0 0 0 0
HWE	**HWE** • Rep 0.98-LocCR0.95 • Rep 0.95-LocCR0.80	 3139 5769	 0 10 (out of HWE in 5 out of 6 populations)	 174 174	 0 0
Outlier loci (identified by both Outflank and PCAdapt)	**Outlier loci** • Rep 0.98-LocCR0.95 • Rep 0.95-LocCR0.80	 3127 5754	 12 15	 174 174	 0 0

Table 4. Summary of quality filtering scenarios with relative number of loci and number of individuals retained in each scenario. P/F indicates whether the scenario is a preliminary or final dataset.

Datasets	Filtering scenario	n Loci	n Ind	P/F
1	Rep0.98_rDepth5_125_IndCR0.68_noMono_LocCR0.95_MAF0.04_2nd	3139	174	P
2	Rep0.98_rDepth5_125_IndCR0.68_noMono_LocCR0.80_MAF0.04_2nd	4584	174	P
3	Rep0.95_rDepth5_125_IndCR0.68_noMono_LocCR0.95_MAF0.04_2nd	3915	174	P
4	Rep0.95_rDepth5_125_IndCR0.68_noMono_LocCR0.80_MAF0.04_2nd	5779	174	P
5	Rep0.98_rDepth5_125_IndCR0.68_noMono_LocCR0.95_MAF0.04_2nd_noOutl	3127	174	F
6	Rep0.95_rDepth5_125_IndCR0.68_noMono_LocCR0.80_MAF0.04_2nd_ noHWE_noOutl	5754	174	F

Population structure analyses

Analyses for population structure were only conducted on the two final datasets: the most stringent and the most relaxed, datasets 5 and 6 (Table 4). These consisted of a final PCA, pairwise population F_{STs} and genetic Admixture analyses.

Furthermore, four additional analyses were conducted on the stringent final dataset (Table 4, dataset 5): spatial autocorrelation (package dartR.spatial), isolation by distance (through a Mantel test, package dartR.spatial), AMOVA (package poppr, Kamvar et al., 2014) and k-means clustering (package adegenet::find.clusters). The results of the spatial autocorrelation, isolation by distance and k-means clustering are reported in Appendix 2.

Principal Component Analysis

The PCA analyses showed a similar clustering pattern as the PCA on the pre-filtered dataset and the preliminary datasets, with Lucinda and Mackay clustering together and separately from all other sampling groups along the primary axis of the PCA plot which explained about 1.1% of variation (Figure 2).

Pairwise F_{STs} among sampling groups

To validate the genetic separation between the Lucinda-Mackay cluster and the Rockhampton-Hervey Bay-Fraser cluster, pairwise F_{STs} analyses were conducted across all sampling groups to test for significance and magnitude of genetic differentiation among sampling groups.
Pairwise F_{STs} were small (0.005-0.006) but significant ($p<0.001$) between Lucinda and Rockhampton Offshore sampling groups (S29, S30), Fraser Offshore and Hervey Bay, and between Mackay and Rockhampton Offshore sampling groups (S29, S30), Fraser Offshore and Hervey Bay (Figure 3).

Admixture analysis

For both final datasets (5 and 6), the cross-entropy criterion calculated by the SNMF algorithm showed the lowest value for K=1, indicating the presence of a single population (Figure 4). The differentiation between the clusters detected through PCA and F_{STs} seems too small to be detected by the Admixture SNMF algorithm. However, given that both PCA and pairwise F_{STs} showed two clusters of sampling groups, the presence of two potential populations was also tested (K=2). The admixture plot for K=2 showed Lucinda-Mackay separating from the southern sampling sites in both datasets (Figure 5, admixture plot for dataset 5; Figure 3 in Appendix 2, for admixture plot for dataset 6). These two populations show different genetic composition (different allele frequencies): both Lucinda and Mackay individuals show a higher percentage of one of the two estimated genetic pools (K1> 58%) (except for 5146SCA23 from Lucinda, K1 = 49%); while all individuals from Rockhampton offshore, Hervey Bay and Fraser Offshore show lower percentages for this genetic pool (K1<58%). The only exceptions are two individuals, one from Fraser Offshore (5014SCA23 K1=73%) and one from Hervey Bay (5172SCA23 K1=62%) (values represent those for dataset 5, but similar values were obtained in dataset 6). Testing for three potential stocks (K=3) did not reveal any additional differentiation (two clusters: Lucinda-Mackay and Rockhampton-Hervey Bay-Fraser) (Figure 2, Appendix 2).

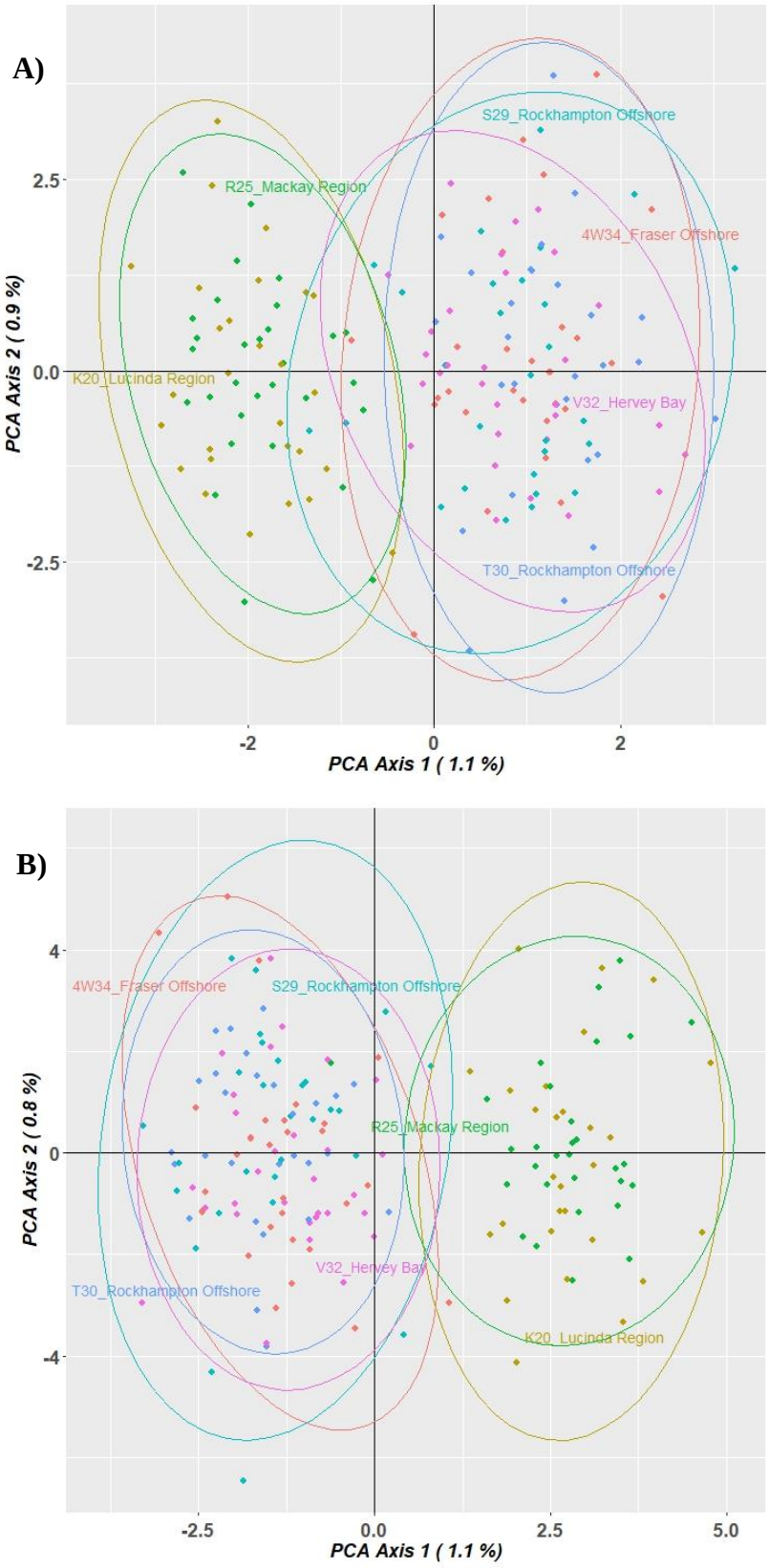

Figure 2. PCA plots for datasets: A) dataset 5 (stringent filtering), B) dataset 6 (relaxed filtering).

AMOVA

The AMOVA analysis on the final stringent filtering scenario (dataset 5) was conducted by defining three levels of clustering: i) individuals, ii) populations, iii) the two clusters identified with both PCA and pairwise F_{STs}, that is the Mackay-Lucinda cluster and the Rockhampton-Hervey Bay-Fraser cluster. This was done to test whether variance between clusters was significantly greater than variance between populations within clusters.

The AMOVA results showed that there is small but significant differentiation between the two clusters (regardless of the genetic distance used) (AMOVA with Nei's genetic distance: F_{CT} = 0.02, p=0.0), while populations within the clusters are not significantly differentiated (AMOVA with Nei's genetic distance F_{SC} = -0.001, p=0.88, Table 5) (see Table 1, Appendix 2 for results with Euclidean distance).

Extremely small variance components for clusters and populations compared to the variance among individuals within populations (i.e. error), would suggest a panmictic population. However, the variation among clusters is significant (F_{CT} = 0.02, p<0.001), while the variation among populations within clusters is not significant (F_{SC} = -0.001, p=0.88). The lack of significant variation among populations within clusters is also indicated by the negative value of the variance (σ^2 = -0.053). As these values represent covariances, they can be negative and this normally indicates lack of genetic differentiation (σ^2 ~ 0, Schneider et al., 2000).

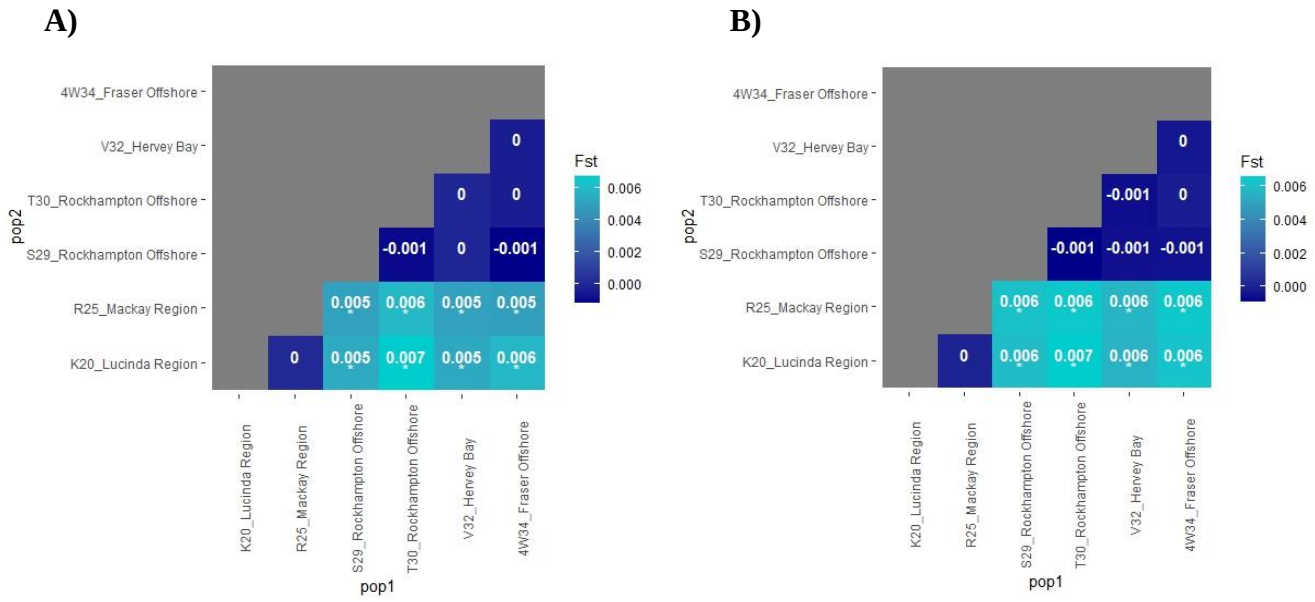

Figure 3. Pairwise population F_{STs} heatmap plots for the two final datasets: A) dataset 5 (stringent filtering), B) dataset 6 (relaxed filtering). Significance values were obtained through 1000 bootstraps and are indicated by an asterixis, p<0.001 (*). F_{STs} values (rounded at 3 decimals) are indicated in each cell representing pairwise comparisons.

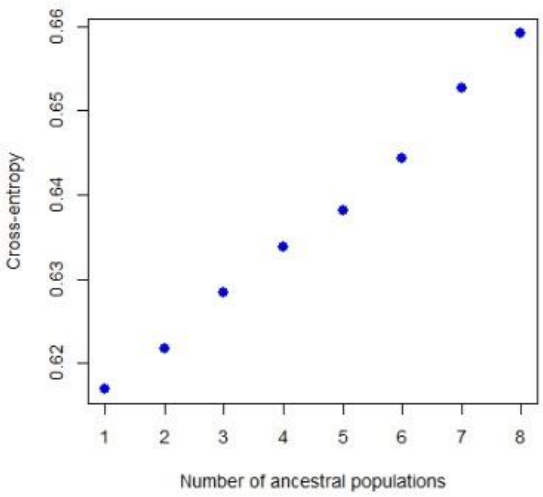

Figure 4. Admixture analysis for dataset 5. Plot of the number of ancestral populations (K) versus the cross-entropy (CE) criterion (defining the number of estimated ancestral populations representing the best fit to the data). The optimal number of K is the lowest CE value.

Figure 5. Admixture analysis for dataset 5. Genetic Admixture plot for K=2 for dataset 5.

Table 5 – AMOVA results using Nei's distance. Indicated are the sums of square deviations (SSD), mean square deviations (MSD), the number of degrees of freedom (df), the variance components and respective percentage, F_{CT} = variance among groups relative to the total variance, F_{SC} = variance among subpopulations within the groups relative to the total variance, F_{ST} = variance among populations relative to the total variance.

	d.f.	SSD	MSD	Variance	% Variance		
Among clusters	1	0.035	0.035	2.9689e-04	2.18	F_{CT} = 0.02	p= 0.0
Among populations	4	0.051	0.012	-1.8593e-05	-0.13 (0)	F_{SC} = -0.001	p=0.88
Among individuals	168	2.235	0.013	1.3308e-02	97.95	F_{ST} = 0.02	
Total	173	2.322	0.013				

Discussion

These analyses confirm genetic differentiation between Ballot's saucer scallops within the Central trawl management region (Lucinda, Mackay samples) and the southern inshore and southern offshore trawl management regions (Rockhampton Offshore, Hervey Bay and Fraser Offshore samples). Despite the genetic differences being small (F_{STs} ~ 0.005), they are significant ($p<0.001$), and are supported by PCA, AMOVA and Admixture analyses. Admixture analyses did not find a significant difference between the two clusters, but the genetic differentiation between the clusters was evident from the plot when two populations were estimated. In addition, spatial autocorrelation analysis indicated that individuals from the same site are more closely related to each other (significantly positive autocorrelation coefficient in the first distance bin), suggesting some level of self-recruitment. However, this analysis is partially biased and lacks power given that geographic coordinates for the samples were only available at the sampling group level.

Considerable levels of self-seeding of Ballot's saucer scallops were reported by Courtney et al. (2015) in a study of the oceanographic influences on reef fish and Ballot's saucer scallops. Larval advection was modelled for the southern inshore and southern offshore areas and demonstrated connectivity between scallops from Yeppoon to Double Island Point, however the central trawl region was not considered in the study. The results of this study further support connectivity between Ballot's saucer scallops in southern inshore and southern offshore trawl management regions.

Whilst the exact drivers for the differentiation of these Ballot's saucer scallops from those in the central trawl management region are unknown it is likely that seasonal patterns in the East Australia Current play a role. Teske et al (2016) found population structure within *Siphonaria diemenensis* (common rocky shore limpet) stock of southern Australia and concluded near shore oceanographic constraints were limiting larval dispersal and contributed to high self-recruitment.

In contrast, other studies have shown connected populations of crustaceans (*Ranina ranina*, *Thenus australiensis*) and molluscs (*Donax deltoides*) along Australia's east coast (Miller et al 2013, McMillan et al 2024, Schilling et al 2022). According to the findings of these studies, seasonal current patterns and extended spawning seasons were likely key drivers of these results. Ballot's saucer scallops differ from these species given they have a specific spawning period (Austral winter) and temperature dependent larval survivability, therefore larval movement and juvenile settlement will be occurring under different oceanographic conditions.

The findings of this study are different to those of O'Brien et al 2003, despite samples being collected from the same regions. It is possible that the differences between the 2 studies are due to the difference in genetic markers used (2 x mtDNA sequence markers and 7 x microsatellite loci versus 3217-5754 SNP loci). Other studies have found differences in results when comparing microsatellite data to SNP data (Vali et al. 2008, DeFaveri et al. 2013). Vali et al (2008), concluded that microsatellite markers may not accurately reflect underlying genomic diversity, due to the ascertainment bias caused by selecting the most polymorphic markers.

The collection of basic biological data (growth and reproduction) for Ballot's saucer scallops within the central region is recommended. This data would provide baseline information for population models

and allow the fishery to be managed in a contemporary manner. Future research to demonstrate the underlying mechanisms driving the population structure seen on Queensland's east coast could include larval dispersal modelling or a finer scale genetic study.

References

Brown, I, Kirkwood, J, Gaddes, S, Dichmont, C and Ovenden, J 1999, Population dynamics and management of Spanner Crabs *(Ranina ranina)* in southern Queensland, project report QO99010, Queensland Department of Primary Industries, Brisbane.

Courtney, A. J., Spillman, C. M., Lemos, R., Thomas, J., Leigh G. M., and Campbell A. B. (2015). Physical oceanographic influences on Queensland reef fish and scallops. FRDC Project 2013/020, Fisheries Research and Development Corporation.

DAFa (2021). Trawl fishery (central region) harvest strategy: 2021-2026, State of Queensland.

DAFb (2021). Trawl fishery (southern offshore) harvest strategy: 2021-2026, State of Queensland.

DAFc (2021). Trawl fishery (southern inshore region) harvest strategy: 2021-2026, State of Queensland.

DeFaveri, J., Viitaniemi, H., Leder, E., and Merila, J. (2013) Characterizing genic and nongenic molecular markers: comparison of microsatellites and SNPs. Molecular Ecology Resources , 13(3), 377-392.

Dredge, M.C.L. (1981). Reproductive biology of the saucer scallop *Amusium japonicum balloti* (Bernardi) in central Queensland waters. Aust. J. Mar. Freshwater Res, 32, 775-87.

French, S.M. (2023) Stock Assessment of Ballot's saucer scallop *(Ylistrum balloti*) in Queensland, Australia, with data to October 2022. Technical Report. State of Queensland, Brisbane.

Frichot, E., & François, O. (2015). LEA: An R package for landscape and ecological association studies. Methods in Ecology and Evolution, 6(8), 925-929.

Georges, A., Mijangos, L., Patel, H., Aitkens, M., & Gruber, B. (2023). Distances and their visualization in studies of spatial-temporal genetic variation using single nucleotide polymorphisms (SNPs). *bioRxiv*, 2023-03.

Gosselin, T. (2020). radiator: RADseq Data Exploration, Manipulation and Visualization using R. R package version 1.1.9 https://thierrygosselin.github.io/radiator/. doi : 10.5281/zenodo.3687060

Gruber, B., Unmack, P. J., Berry, O. F., & Georges, A. (2018). dartr: An r package to facilitate analysis of SNP data generated from reduced representation genome sequencing. Molecular ecology resources, 18(3), 691-699.

Jombart, T., Kamvar, Z. N., Collins, C., Lustrik, R., Beugin, M. P., Knaus, B. J., & Jombart, M. T. (2018). Package 'adegenet'. *Github repository:< https://github. com/thibautjombart/adegenet*.

Kamvar, Z. N., Tabima, J. F., & Grünwald, N. J. (2014). Poppr: an R package for genetic analysis of populations with clonal, partially clonal, and/or sexual reproduction. *PeerJ*, *2*, e281.

Luu, K., Bazin, E., & Blum, M. G. (2017). pcadapt: an R package to perform genome scans for selection based on principal component analysis. Molecular ecology resources, 17(1), 67-77.

McMillan, M. N., Brunjes, N., Williams, S. M. and Holmes, B. J. (2024) *Broad-scale genetic population connectivity in the Moreton Bay Bug (Thenus australiensis) on Australia's east coast.* Hydrobiologia, 851 . pp. 2347-2355. ISSN 1573-5117

Miller Adam D., Versace Vincent L., Matthews Ty G., Montgomery Steven, Bowie Kate C. (2013). Ocean currents influence the genetic structure of an intertidal mollusc in southeastern Australia – implications for predicting the movement of passive dispersers across a marine biogeographic barrier. Ecology and Evolution 2013; 3(5): 1248–1261

O'Brien, Elizabeth, Bartlett, Jason, Crump, Provan, Dixon, Bryony and Duncan, Peter (2005). Enhancement of Saucer Scallops (*Amusium balloti*) in Queensland and Western Australia—Genetic Considerations, FRDC Project 2003/033, Fisheries Research and Development Corporation.

Robledo-Ruiz, D. A., Austin, L., Amos, J. N., Castrejón-Figueroa, J., Harley, D. K., Magrath, M. J., & Pavlova, A. (2023). Easy-to-use R functions to separate reduced-representation genomic datasets into sex-linked and autosomal loci, and conduct sex assignment. Molecular Ecology Resources. (Early View)

Schilling Hayden T., Hewitt Daniel E., Malan Neil, Taylor Matthew D., and Johnson Daniel D. (2022). Cross-jurisdictional larval supply essential for eastern Australian spanner crabs (Ranina ranina). Marine and Freshwater Research 73(11) 1352-1367.

Schneider, S., Roessli, D., and Excoffier, L. 2000. ARLEQUIN, Version 2.000. A Software for Population Genetics Data Analysis. Genetics and Biometry Laboratory, University of Geneva, Switzerland

Smouse, P. E., & Peakall, R. O. D. (1999). Spatial autocorrelation analysis of individual multiallele and multilocus genetic structure. *Heredity*, *82*(5), 561-573.

Vali, U., Einarsson, A., Waits, L. & Ellegren, H. (2008). To what extent do microsatellite markers reflect genome-wide genetic diversity in natural populations? Molecular Ecology, 17(17), 3808-3817.

Whitlock, M. C., & Lotterhos, K. E. (2015). Reliable detection of loci responsible for local adaptation: Inference of a null model through trimming the distribution of F ST. The American Naturalist, 186(S1), S24-S36.

Wortmann, J. (2021) *Stock Assessment of Ballot's saucer scallop (Ylistrum balloti) in Queensland, Australia.* Technical Report. State of Queensland, Brisbane.

Wortmann, J. (2022) *Stock Assessment of Ballot's saucer scallop (Ylistrum balloti) in Queensland, Australia.* Technical Report. State of Queensland, Brisbane.

Zheng, X., & Zheng, M. X. (2013). Package 'SNPRelate'. A package for parallel computing toolset for relatedness and principal component analysis of SNP data.

Appendix 1

Sampling procedure

1. Thaw shell in a single layer using a basket the day prior to processing. This allows fluid to drain away and possibly avoid cross contamination.
2. Using a scallop measurer, measure the shell height to the nearest 1 mm.
3. Place scallop on petrie dish on the scales (ensure the scales are tared correctly) and weigh the whole scallop to the nearest 0.01 gram.
4. With a filleting knife carefully remove the mantle and separate the adductor muscle from the top shell to keep the whole scallop intact (Figure 1A). Open scallop.
5. Using forceps fold over the mantle and the first layer of gills to expose the gonads.
6. Using surgical scissors and forceps, carefully remove gonads including parts of the digestive tract and weigh the gonad to the nearest 0.01 gram (Figure 1B).

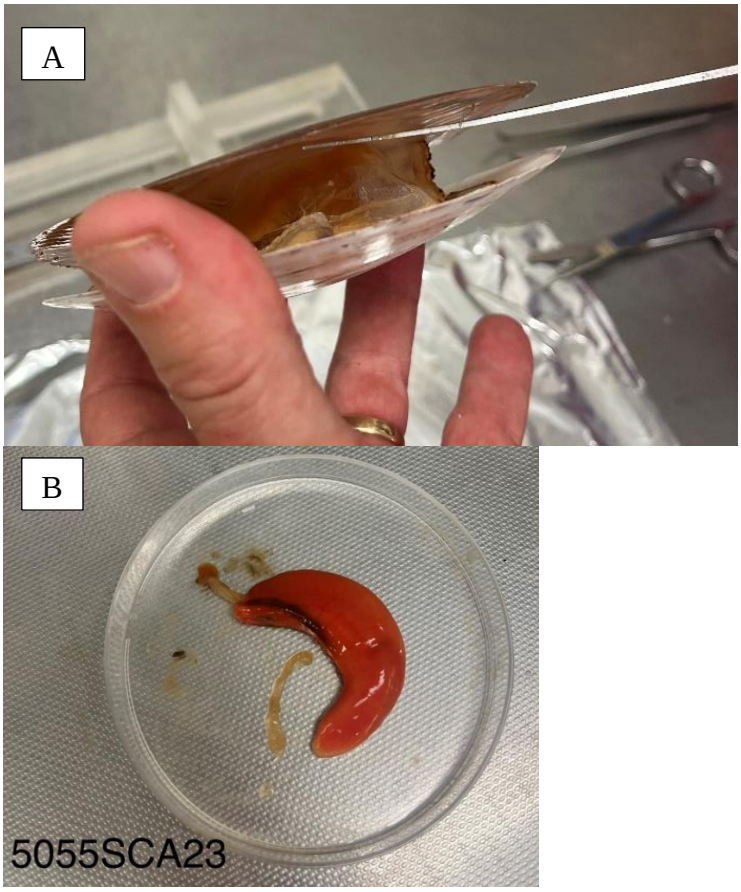

Figure 1. A) Removing the mantle from the top shell to keep whole scallop intact. B). Gonad that has been removed from scallop to be weighed.

7. Assign a macroscopic stage with the use of Table 1 and Figure 2.

Table 1. External Features used to determine gonad development in *A. japonicum balloti*. Sourced from Dredge 1981

Stage	Classification	Description
1	Immature or resting	Gonad colourless, gender not evident. Volume of gonad not greater than volume of digestive loop
2	Initial development	Gonad translucent, gender not evident. Volume of gonad more than twice that of digestive loop
3	Early development	Gender evident. Male gonads white, female gonads pink–red. Gonads less than half volume of fully mature gonads, ventral part of digestive loop visible
4	Late development	Gender evident. Male gonads uniform white, female gonads pink–red. Digestive loop almost entirely enveloped, with only ventral curve visible. Gonads have not reached ultimate volume or turgidity
5	Mature	Gender evident. Male gonads creamy, lustrous, and in some instances follicles are visible to naked eye. Female gonads pink–red, lustrous in fresh material. Gonads envelop digestive loop and are turgid
6	Spent	Gender usually evident, but digestive loop visible through most of gonad. Gonad flaccid, with free water in lumen. Males grey, females pink to light brown

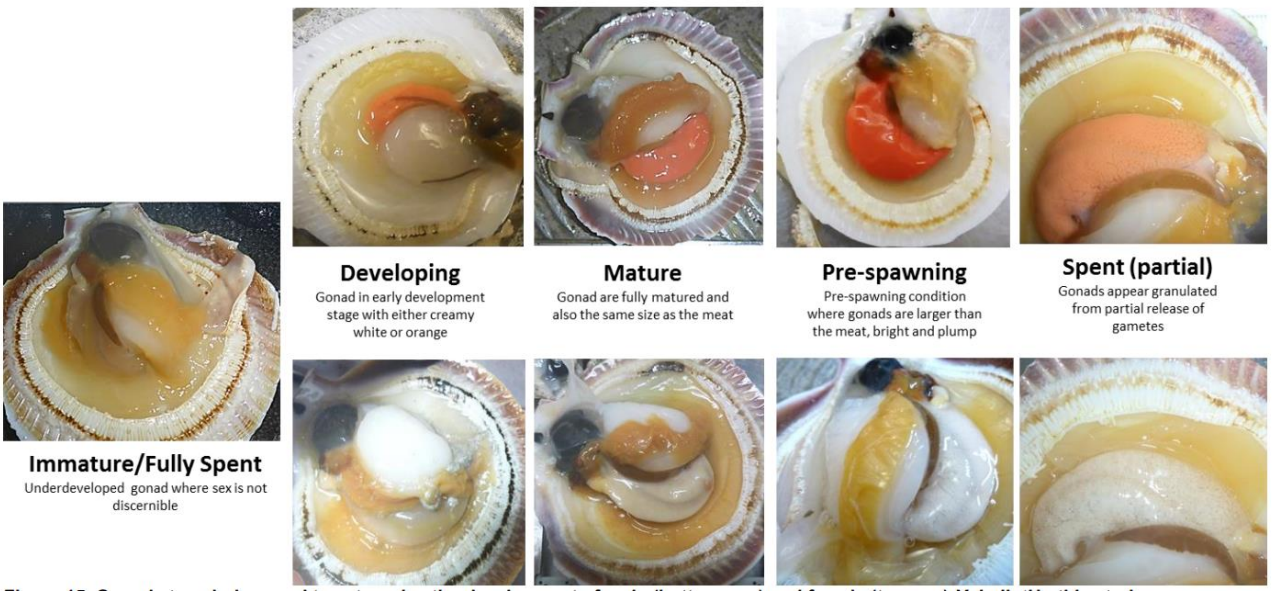

Figure 2. Gonad stage index used to categorise the development of female (top row) and male (bottom row) for *Y. balloti*. Sourced from A. Chandrapavan, M.I. Kangas & N. Caputi 2020

8. Using surgical scissors, carefully cut the adductor muscle out from the rest of the scallop (Figure 3) and weigh to the nearest 0.01 gram. Note- if muscle has high fluid content, dab with some paper towel to remove excess.

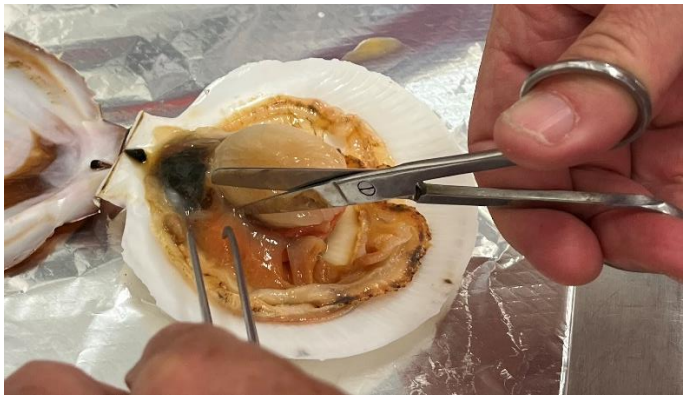

Figure 3. Cutting the scallop from the rest of the content to remove adductor muscle.

9. Place the scallop on the dedicated genetic chopping board.
10. Set up equipment as per Fishery Monitoring procedure for genetic tissue sampling. To avoid cross contamination ensure all utensils are cleaned in the detergent, bleach then water solutions before, during, and after, taking each scallop tissue sample.
11. Using a clean knife cut the adductor muscle in half in one cutting motion.
12. Using a clean scalpel and forceps remove a pea sized piece of muscle from the middle of one side of the cut scallop (Figure 4). Place tisue sample into a 2 ml genetic vial filled with molecular grade ethanol.

Figure 4. Genetic sample removed from the middle of the adductor muscle.

13. Record the unique genetic number for the tissue sample.
14. Using hydrogen peroxide wipes, wipe the aluminium covered chopping board after each scallop is processed.
15. Store trays of genetic sample vials in the freezer at the end of the processing session.

References

M.C.L. Dredge (1981). Reproductive biology of the saucer scallop *Amusium japonicum balloti* (Bernardi) in central Queensland waters. Aust. J. Mar. Freshwater Res, 32, 775-87.

A. Chandrapavan, M.I. Kangas, N. Caputi (2020). Understanding recruitment variation (including the collapse) of Ballot's saucer scallop stocks in Western Australia and assessing the feasibility of assisted recovery measures for improved management in a changing environment, Fisheries Research Report No. 308 Department of Primary Industries and Regional Development, Western Australia. 76pp.

Appendix 2

Principal Component Analysis

A Principal Component Analysis (PCA) conducted on the original dataset (pre-filtering), showed two clusters: one including Lucinda and Mackay, and the other including all other sampling groups (Rockhampton offshore (R29 and S29), Hervey Bay and Fraser Offshore) (Figure 1).

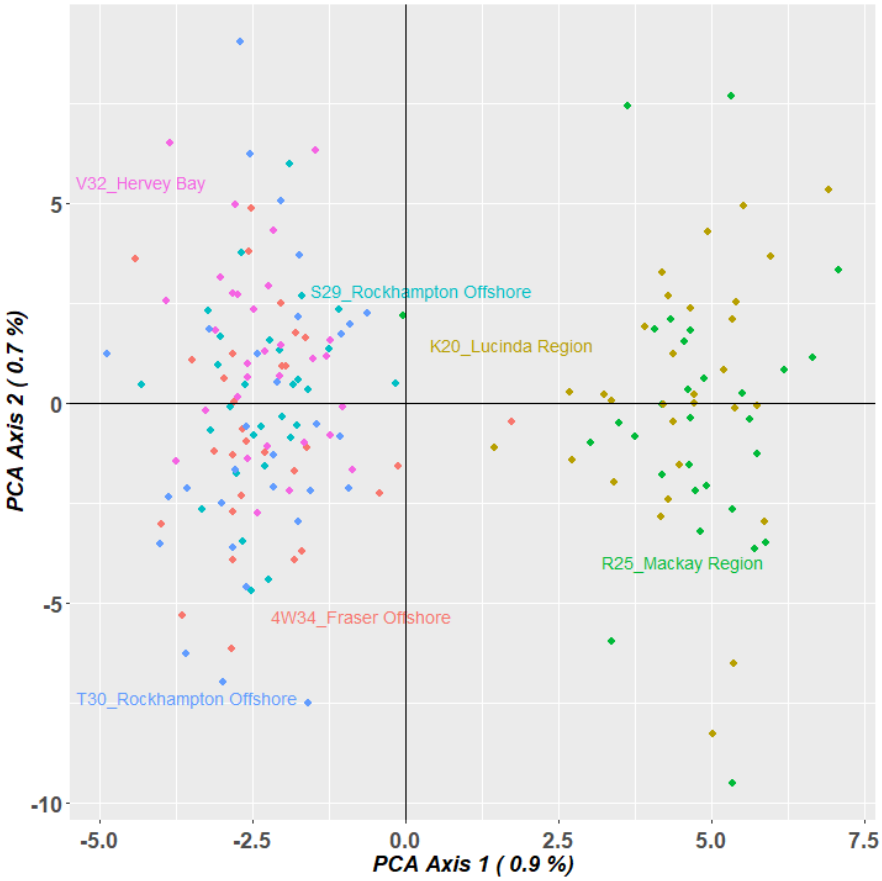

Figure 1. PCA plot conducted on the original dataset pre-filtering. Individuals are labelled by sampling group.

These clusters were separated along the primary axis of the PCA plot explaining about 1% of variation.

Admixture plots for dataset 5 - K3 and for dataset 6 - K2

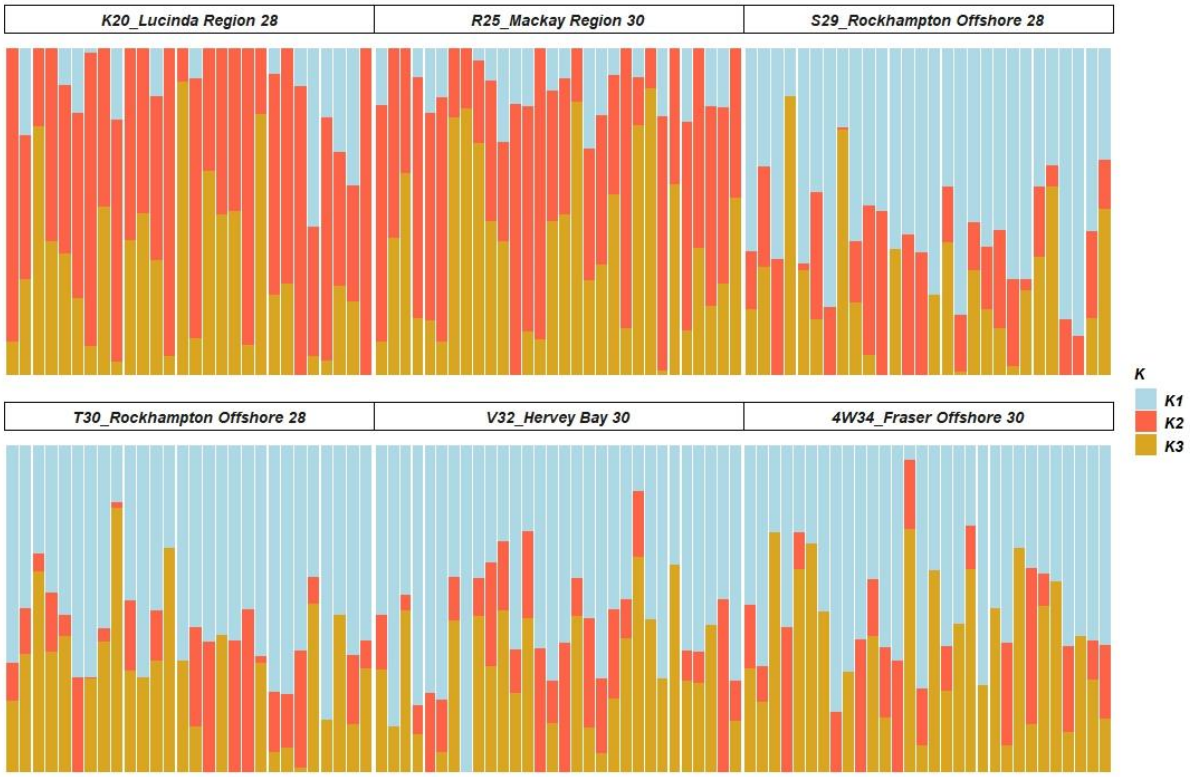

Figure 2. Genetic Admixture plot for K=3 for dataset 5.

Figure 3. Genetic Admixture plot for K=2 for dataset 6.

AMOVA with Euclidean distance

Table 1. AMOVA results using Euclidean distance.

	d.f.	SSD	MSD	Variance	% Variance		
Among clusters	1	1500.40	1500.40	8.63	1.00	F_{CT} = 0.010	p= 0.0
Among populations	4	3329.22	832.30	-0.053	-0.06 (0)	F_{SC} = -0.0006	p=0.87
Among individuals	168	142431.37	847.80	847.80	99.05	F_{ST} = 0.009	
Total	173	147260.99	851.21				

Additional Analysis on dataset 5 (stringent)

Spatial Autocorrelation

Method

Spatial autocorrelation assesses whether individuals which are closer geographically are also more similar genetically by calculating a correlation coefficient between geographic and genetic distance. The autocorrelation coefficient "r" is calculated for each pair of individuals in each specified distance class (or "bins") (Smouse and Peakall, 1999). This method can provide information on the dispersal range of the target species, as individuals will be more genetically similar within their dispersal range.

In this approach, significance testing is achieved through two methods: first 95% bootstrap confidence intervals are calculated around the value of r for each bin by drawing with replacement from within the set of relevant pairwise comparisons for a given distance class; second a one-tail permutation test is conducted to calculate the null distribution and the 95% confidence intervals around the null hypothesis.
In the first method, if bootstrap confidence intervals do not overlap zero, fine-scale spatial genetic structure is present. In the second method, if the autocorrelation coefficient "r" falls outside of this range, significant fine-scale spatial genetic structure is present.
Spatial autocorrelation (package dartR.spatial::gl.spatial.autoCorr) was assessed only for the final dataset, the most stringent filtering scenario (Table 6, dataset 5).

Results

Spatial autocorrelation was conducted on dataset 5 (table 6) with a distance bin of 100 km and 100 permutations (Figure 4). As geographic coordinates are available only at the sampling group level, spatial autocorrelation cannot be estimated within sampling locations; therefore, all sampling groups were pooled in a single population for this analysis. Figure 6 shows that spatial autocorrelation is positive and significant up to 200km (confidence intervals calculated through bootstraps are above zero, and they do not overlap with the null distribution (red area)). There isa positive spatial autocorrelation within the 500 km distance class. This could be due to the higher genetic similarity between the Mackay and Lucinda sampling groups, as demonstrated through PCA and pairwise F_{ST}s. When spatial autocorrelation was conducted on the same dataset but without the Lucinda sampling group, the positive r value at 500 km disappeared (Figure 5). Removing the Lucinda sampling group, showed that spatial autocorrelation is positive and significant only within the 100 km distance bin. This could suggest that the dispersal ability of scallop larvae is in the order of 100 km. However, it must be noted that since geographic coordinates are available only at the level of the sampling group (individuals within a sampling group have the same geographic coordinate), this analysis has minimal statistical power. Furthermore, this could bias the results of this analyses as well as limiting its inference power.
Spatial autocorrelation was also conducted on dataset 5 to compare differences in dispersal among sex, but no differences could be detected with the data available (Figure 6).

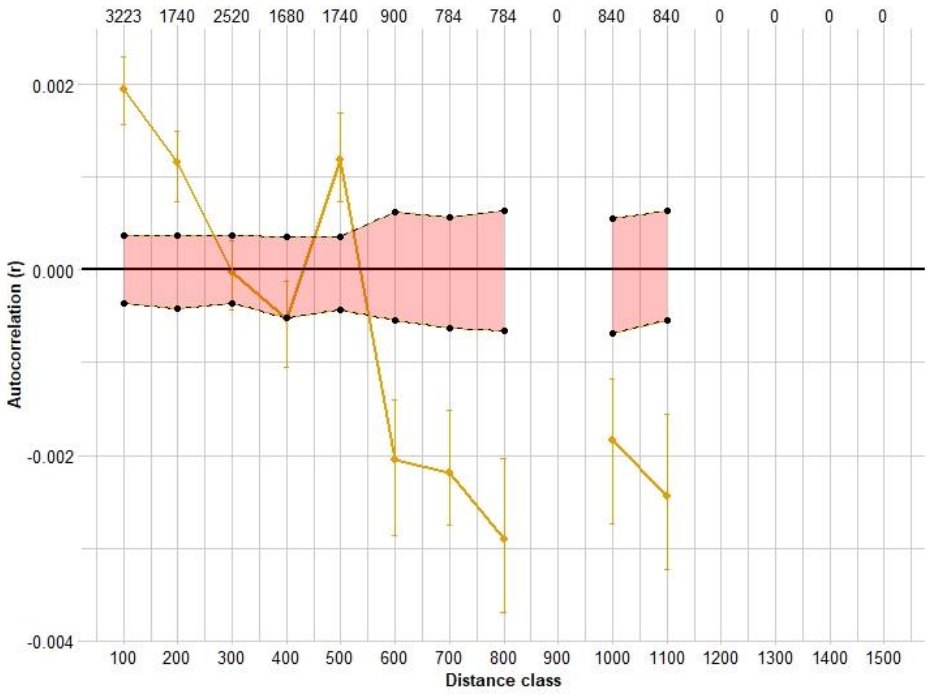

Figure 4. Spatial autocorrelogram plot for dataset 5, with 100 km distance bins, and 100 repetitions for both bootstraps and permutations. The bars represent confidence intervals around the r value in each bin obtained through bootstraps, while the red area represent the r values of the null distribution obtained through permutations. The gap in the r value line indicated that no pairwise comparisons where available in that distance bin class.

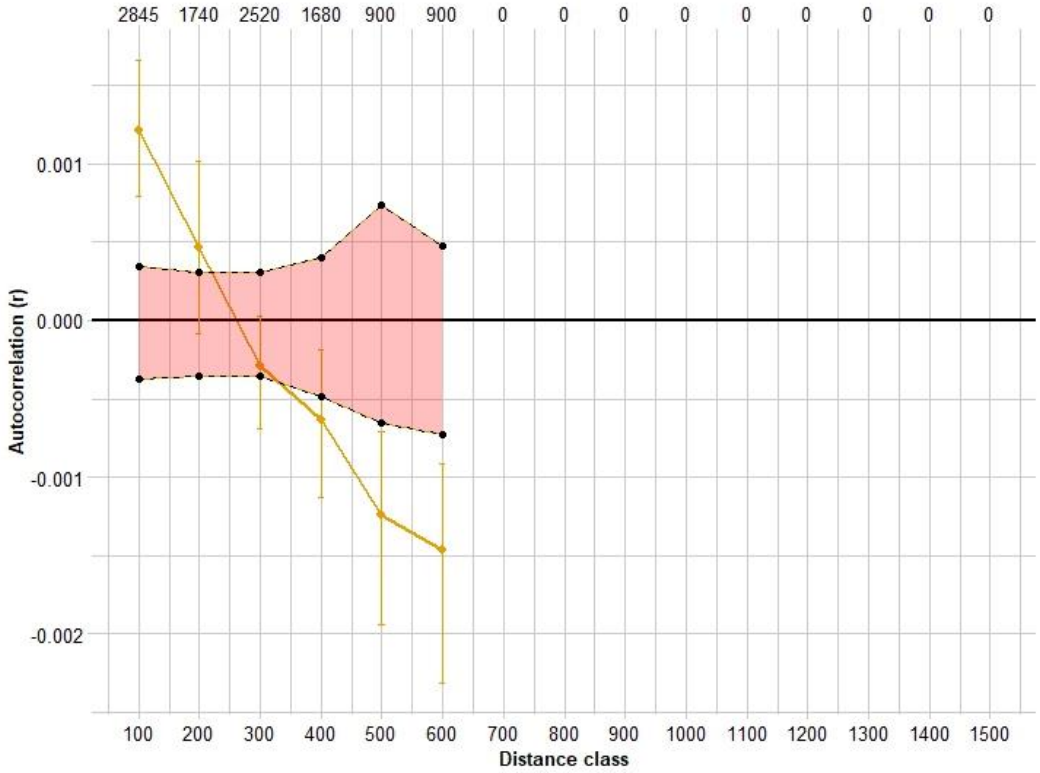

Figure 5. Spatial autocorrelogram plot for dataset 5 without Lucinda, with 100 km distance bins, and 100 repetitions for both bootstraps and permutations. The bars represent confidence intervals around

the r value in each bin obtained through bootstraps, while the red area represent the r values of the null distribution obtained through permutations.

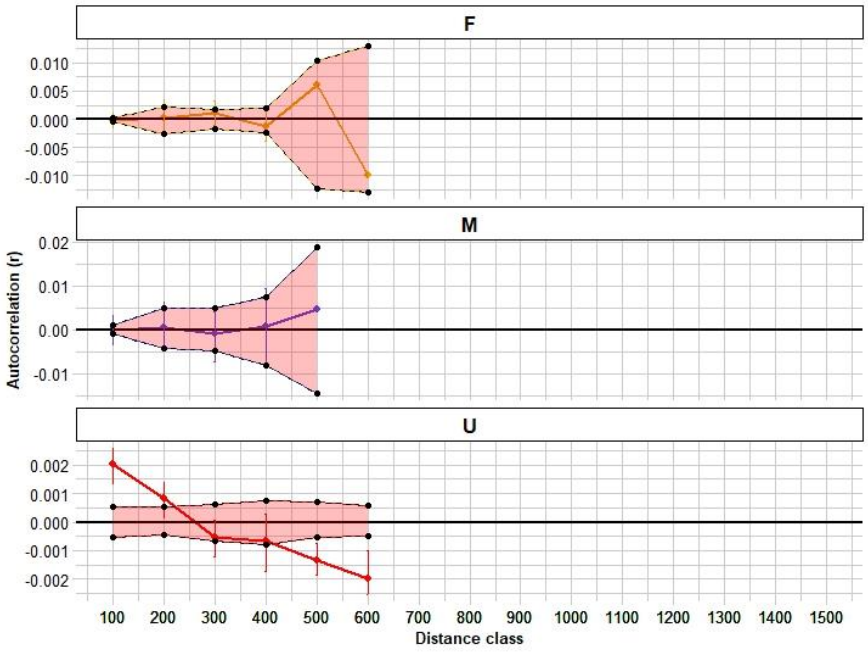

Figure 6. Spatial autocorrelogram plot for dataset 5 with comparison among sex, 100 km distance bins, and 100 repetitions for both bootstraps and permutations.

Isolation by distance

Method

Isolation by distance was conducted with dartR (gl.ibd), using population F_{st} as genetic distance, Euclidean distance between geographic coordinates, and 999 permutations to assess statistical significance.

Results

Isolation by distance was estimated on dataset 5 through a Mantel test with population F_{st} used as the measure of genetic distance. There was a significant positive but weak correlation between geographic and genetic distance throughout the sampling range when Lucinda was retained in the dataset (Figure 7) ($r^2 = 0.48$, $p=0.02$), however the correlation was not significant when Lucinda was removed ($r^2 = 0.36$, $p=0.17$). The results indicate absence of isolation by distance as the linear relationship between geographic distance and genetic distance did not adequately fit the data.

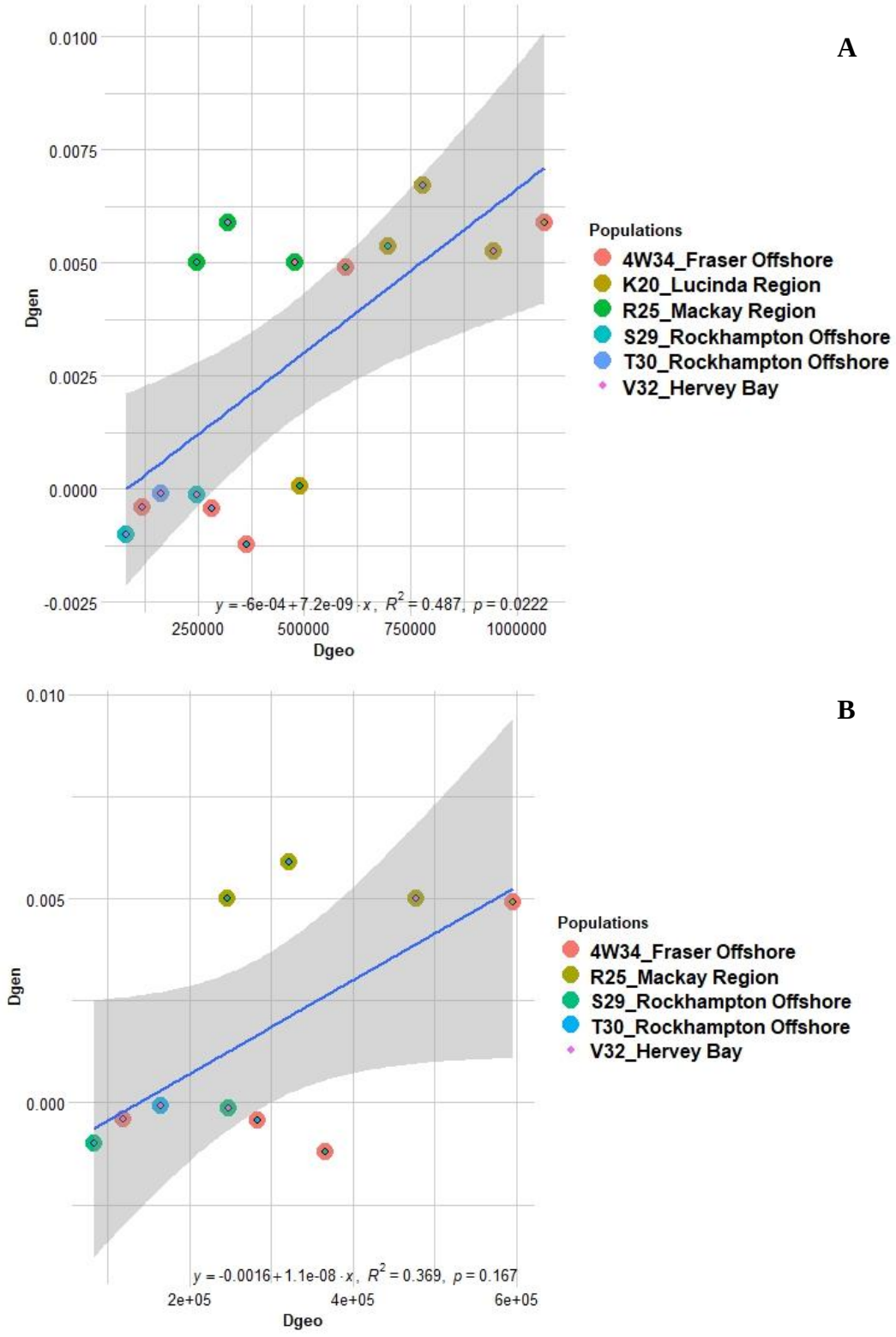

Figure 7. Mantel test for Isolation by Distance for dataset 5, a) with all sampling groups, b) without Lucinda. The colours around each point represent which two populations that pairwise comparison relates to.

K-means clustering for dataset 5

K-means clustering was conducted with the R package adegenet (find.clusters), setting the maximum number of clusters to 8 (greater than the number of sampling groups) and the statistic to BIC.
K-means clustering detected a single cluster of individuals (K=1), similarly to the SNMF Admixture analysis (Figure 8, left). When 2 clusters were requested to be defined in the data (K=2), the k-means clustering algorithm classified all individuals from Lucinda and Mackay in a single cluster (I) and all individuals from the other sampling groups (Rockhampton offshore S29, S30, Hervey Bay and Fraser offshore) in a separate cluster (II). However, four individuals were identified as belonging to the Mackay-Lucinda cluster even if they originally derive from other regions: 5101SCA23, 5092SCA23 and 5106SCA23 from Rockhampton offshore (S29) and 5014SCA23 from Fraser offshore (Figure 8, right). These results are like the Admixture results where 2 individuals one from Hervey Bay and one from Fraser (the same one identified here) were detected as having allele frequencies more similar to the Lucinda-Mackay group.

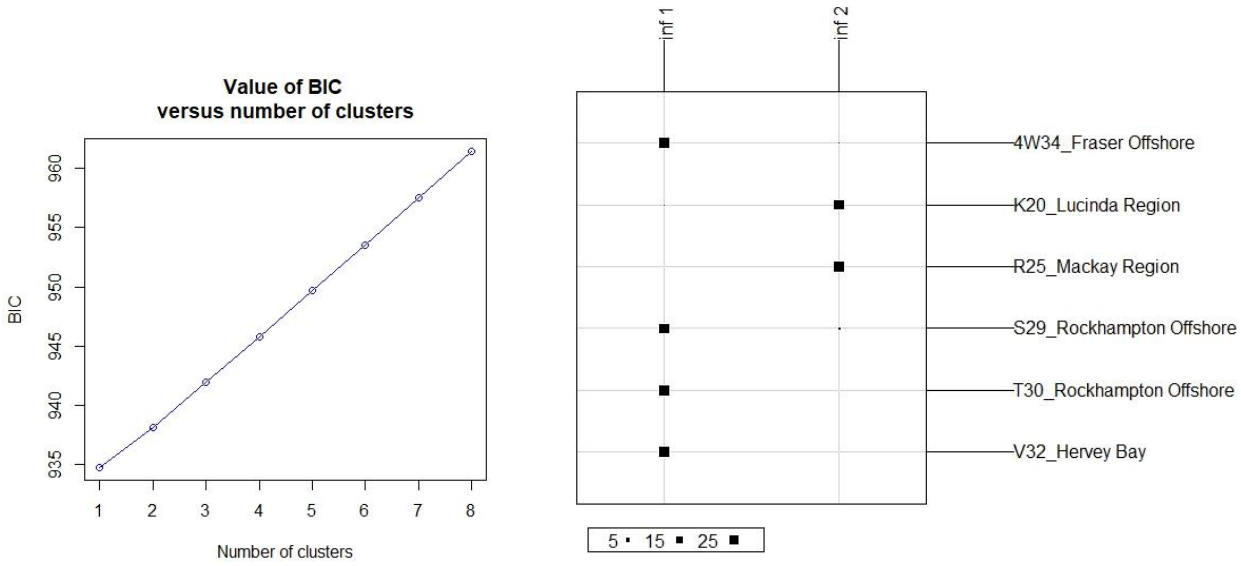

Figure 8. K-means clustering for the final stringent dataset (dataset 5). Left: Number of clusters (K) plotted against the BIC factor. The optimal number of clusters is the one with the lowest BIC, in this case K=1. Right: Groups obtained with K-means clustering when 2 clusters were requested to be defined (K=2), with comparisons of clusters identified by k-means clustering with the original sampling groups.