PDRS: Use of next generation DNA technologies for revealing the genetic impact of fisheries restocking and ranching

Jonathan Sandoval-Castillo, Nick Robinson, Lachlan Strain, Anthony Hart and Luciano B. Beheregaray

Project No. 2012/714

May 2015

This project was instigated and carried out by Flinders University, Sturt Road, Bedford Park, SA 5042, Australia and Nofima, PO Box 5010, 1432 Ås, Norway and the Department of Fisheries Western Australia, PO Box 20, North Beach WA 6025

The Australian Seafood CRC is established and supported under the Australian Government's Cooperative Research Centres Program. Other investors in the CRC are the Fisheries Research and Development Corporation, Seafood CRC company members, and supporting participants.

Office Mark Oliphant Building, Laffer Drive, Bedford Park SA 5042
Postal Box 26, Mark Oliphant Building, Laffer Drive, Bedford Park SA 5042
Tollfree 1300 732 213 Phone 08 8201 7650 Facsimile 08 8201 7659
Website www.seafoodcrc.com ABN 51 126 074 048

ISBN: 978-0-9756044-9-6

Table of Contents

Non-Technical Summary

2012/714 PDRS: Use of next generation DNA technologies for revealing the genetic impact of fisheries restocking and ranching

PRINCIPAL INVESTIGATOR: Prof Luciano B. Beheregaray, luciano.beheregaray@flinders.edu.au

ADDRESS: Flinders University, Sturt Road, Bedford Park, SA 5042, Australia

PROJECT OBJECTIVES:

- To develop new expertise and transfer genomic methods (i.e. Genotype by Sequencing, GBS) that are currently mostly used for commercially important species of the northern Hemisphere, to benefit the fisheries or aquaculture of key commercial species in Australia;
- To generate genomic data for Australian abalones using a GBS approach known as RAD-seq to analyse changes and patterns in neutral and functionally relevant genetic diversity (i.e. variation related to fitness) associated with restocking and ranching and advise on management strategies for the species concerned;
- Through objective 2, to assist in creating genetically healthy/sustainable restocked populations;
- Through objective 2, to assist in improving returns from ranching while having minimal impact on natural populations;
- Through objective 2, to assist in maintaining healthy fisheries;
- To develop expertise in genetic management associated with enhancement and related interventions.

PROJECT SUMMARY

The project developed new diagnostic genomic tools to study natural population genetic structure and to monitor the success of two restocking operations for commercial abalone fisheries in Western Australia (WA). The Roe's abalone (*Haliotis roei*) re-stocking project is being undertaken as a response to a catastrophic mortality event that occurred during the summer of 2010/11 in WA caused by a 'marine heatwave'. The Greenlip abalone (*Haliotis laevigata*) project examined commercial scale stock enhancement as a fisheries management strategy in an attempt to increase the economic yield of the fishery. We collected 420 samples of Roe's abalone (*H. roei*) from eight localities and 372 samples of Greenlip abalone (*H. laevigata*) from thirteen localities, covering a large range of the distribution of both species. All 792 samples were analysed using the new tools, producing a total of 31,008 high quality genomic markers for Roe's abalone and 69,720 high quality genomic markers for Greenlip abalone. These markers are now available in the form of SNPs (single nucleotide polymorphisms) and represent the largest sets of DNA markers to date for Australian abalones. This powerful resource of genome-wide markers can now be used in a large number of applications. This includes but is not limited to association studies looking for traits underpinning performance in aquaculture and restocking, adaptation of captive and natural populations to temperature changes, genome mapping of phenotypic traits under selection in aquaculture, and to understand how both natural selection and the aquaculture environment influence the ability of populations to maintain genetic diversity and to respond to changing conditions. The screening of genome-wide variation in samples collected from the wild show that 'neutral' SNPs (i.e. DNA markers that are not under the influence of natural selection) support the existence of one single abalone population with high connectivity across the geographic range sampled. This result was obtained for both Roe's and Greenlip abalones. However, when we considered the SNP markers under natural selection, we found three genetically distinct groups of populations for Roe's abalone and five genetically distinct groups of populations for Greenlip abalone. These are refereed herein as 'adaptive groups' of abalones. We further conducted a range of statistical analyses that disclosed significant associations between the distribution of these adaptive groups and the spatial variation of key environmental parameters, including differences in temperature and oceanographic variables. These results are critically important for re-stocking programs of Roe's abalone in their northernmost distribution and for stock enhancement of Greenlip abalone in the main fishing areas. The project has provided an outstanding resource and detailed base knowledge that will assist the management of abalone fisheries and aquaculture in Australia. First, for each species thousands of DNA markers were identified and characterised; these markers will be useful for monitoring the genetic health of stocks in both species. Second, high genetic connectivity was detected across the sampling area, but more than one adaptive group were detected for each species; these groups are considered here as 'locally adapated' and this finding will help managers choose which abalone source populations are likely to perform best in specific environments (ie. likely fitness) which will improve the chances of success of specific restocking and stock enhancement programs. Third, we have developed new capabilities and expertise regarding the application of genomic approaches to improve the fishery and aquaculture of abalone in Australia; this knowledge and expertise can be applied to the management of other important fishery and aquaculture species in Australia.

OUTCOMES ACHIEVED:

Roe's abalone

- The *Haliotis roei* restocking being conducted at the Kalbarri Cliffs has successfully translocated adult Roe's abalone (>50 mm shell length) to the study sites at "effective" population sizes. Unfortunately juvenile (0+ age) abalone (recruits) have not yet been detected and so the emphasis of the project's genetic component was changed to assessing and comparing the genomic diversity and differentiation between the remnant Kalbarri Cliff population and other *H. roei* populations around the Australian coast. Our advice now incorporates the identification of source populations that would be suitable for restocking the Kalbarri Cliffs area.
- Levels of genome-wide diversity in *H. roei* were similar in all populations analysed. The highest diversity was detected in the populations close to Perth.
  - Emphasis should nonetheless be placed on maximizing the genetic diversity of restocking efforts by using high founder population numbers (e.g. to give genetically effective population sizes, N_e, greater than 500).
- The level of genetic diversity within the remnant Kalbarri Cliff population is comparable to that found within other *H. roei* populations around the coast (e.g. slightly higher than genetic diversity found within the neighbouring Lucky Bay population).
  - Because the sustained period of elevated water temperatures was a relatively recent event, there has not been time for loss of genetic diversity to occur from the remnant population. However, due to the low census and small effective population size of the remnant population we expect that loss of genetic variation would occur very rapidly in the next generations, unless there is significant natural recruitment of abalone from neighbouring populations or extensive restocking programs are implemented to enhance recruitment.
- If we consider only the fraction of the genomic dataset that is **not** under natural selection, the resulting pattern shows high connectivity and low population differentiation across the wide region sampled (i.e. from Kalbarri Cliff to Spencer Gulf). This is similar to patterns of connectivity reported for Australian abalones based on traditional genetic methods ([e.g. 1]) that are unable to pick up signs of selection.
- If we consider the section of the genome under selection (i.e. functionally relevant section of the genome), a minimum of three genetically distinct groups can be clearly defined in our dataset. These are: 1. Kalbarri Cliff to Lucky Bay, 2. Greenough to Augusta, and 3. Albany to Spencer Gulf populations. Thus, we can conclude that different natural selective pressures across the range of the species have contributed significantly to the overall pattern of differentiation detected.
- More will be deduced about the actual forces of natural selection on these populations, and the genetic processes effected, once the patterns of differentiation at outlier markers (i.e. DNA markers under selection) are correlated with the geospatial data, and when some of these markers are annotated with gene identities (in progress).
  - Restocking programs should focus primarily on sourcing animals from the close geographical location of Lucky Bay (if possible), as this location is more likely to share similar environmental conditions to the collapsed population and as these individuals are more likely to contain useful adaptations promoting survival in the restocked environment. However, if *H. roei* numbers are low in Lucky Bay and it is not possible/desirable to source animals from this area, then sourcing animals from other areas further down the coast must be considered.
- The sequencing technique used to genotype *H. roei* was found to be very effective for Greenlip abalone (*Haliotis laevigata*) also.

Greenlip abalone

- Levels of genome-wide diversity were similar in all populations analysed and comparable to those detected for Roe's populations. The highest diversity was detected in the eastern populations.

o The use of large numbers of broodstock is recommended to maintain the genetic variability within and between populations.

- If we consider only the fraction of the genomic dataset that is **not** under natural selection, the analyses indicate high population connectivity across the sampled area. This is similar to the pattern reported for Roe's abalone.
- When only the section of the genome under selection is considered, five genetically distinct groups can be clearly defined in our dataset. These are: 1. Outback to Windy Outside, 2. Parrys Bay and Whalebone Point, 3. Inner Island to Masons, 4. Fanny Cove to Burton Rocks, and 5. Rob Island to Gulch populations.The seascape analysis supports the existence of four oceanographic regions, which are partially congruent with the five adaptive clusters in *H. laevigata*. The adaptive differentiation between genetic clusters is significantly correlated with geographic variation in oxygen concentration.

o The stock enhancement program should use broodstock taken directly from the genetic group to be enhanced. If broodstock cannot be taken from the genetic group to be enhanced, they should be taken from the most genetically similar group available. Since the genetic differentiation detected is probably adaptive, and is therefore expected to result in varying degrees of fitness/performance in the destination environment, the stock enhancment program should be genetically tested to monitor the contribution of different broodstock to the next generation.

LIST OF OUTPUTS PRODUCED:

- All 792 samples were analysed using the new tools, producing a total of 31,008 high quality genomic markers for Roe's abalone and 69,720 high quality genomic markers for Greenlip abalone.
- These markers are now available in the form of SNPs (single nucleotide polymorphisms) and represent the largest sets of DNA markers to date for Australian abalones.
- Comprehensive genomic analysis of genetic diversity and connectivity in *H. roei* stocks in Western and southern Australia.
- Comprehensive genomic analysis of genetic diversity and connectivity in *H. laevigata* stocks in Western Australia.
- Recommendations for capturing genetic diversity and adapted genotypes for increasing the chances of *H. roei* restocking and *H. laevigata* stock enhancement success.
- Development of previously non-existent capacity in genomic profiling, next-generation sequencing and bioinformatics analyses for the Australian Seafood CRC.

1. Introduction and Background

Restocking has long been practiced as a means of enhancing fisheries, conserving natural populations and reintroduction. But only in the last few decades has attention been given to the genetic effects of the restocking process, and has genetics been used as a tool to monitor the success of restocking. The first step for such genetic studies is normally to develop a set of marker loci (commonly a set of microsatellites or single nucleotide polymorphisms, SNPs) for each species under study. This can be a lengthy and expensive process in itself, and employs the use of DNA sequencing. Once marker loci have been developed for the species in question, they need to be tested and used to genotype a sample of animals from the populations under study. The cost of genotyping limits the number of loci and animals that can be genotyped for the study, and hence the power of the studies' ability to discriminate effects.

In the last few years a major leap in our ability to sequence whole genomes has occurred. Sequencing costs have reduced, accuracy and throughput has improved, and our ability to analyse the tera bytes of data produced from a single sequencing run have greatly improved. One powerful technique that has recently emerged has become generally known as Genotyping-By-Sequencing (GBS). GBS uses next generation sequencing where the DNA of individuals are bar-coded such that the SNP genotypes of these individuals at many different loci can be called direct from the sequence data. Variants of GBS include restriction associated DNA sequencing or RAD-seq [2] and double digest RAD-seq [3]. Both these techniques involve producing a “reduced representation” DNA library from the genome for genotyping by sequencing purposes [4]. RAD-seq employs restriction enzymes to reduce genome complexity, which greatly simplifies analysis for species with high levels of genetic diversity. There is no requirement for marker development for specific species, and the "genotyping" and sequencing is performed in a single step. GBS is proving to be quick, extremely specific and highly reproducible, and is finding applications in population genetics, phylogenetics and quantitative genetics [eg. 5; 6; 7], including application to commercially important species in the northern Hemisphere such as crops and salmonid fishes [eg. 8].

In this project we develop GBS (RAD-seq) as a diagnostic tool to study natural population genetic diversity and structure to assist in examining restocking and stock enhancement operations being trialled for commercial abalone fisheries in WA. The Roe's abalone *(Haliotis roei)* restocking project is being undertaken as a response to a catastrophic mortality event that occurred during the summer of 2010/11 in WA caused by a "marine heatwave". This project arose following the complete closure of the fishery north of Moore River (Area 8) to protect any remaining stock and from the desire by industry to examine the possibility of assisted recovery (Seafood CRC 2011/762: Recovering a collapsed abalone stock through translocation). The stock enhancement project examined commercial scale stock enhancement as a fisheries management strategy in an attempt to increase the economic yield of the Greenlip abalone (*Haliotis laevigata*) fishery (Seafood CRC 2009/710: Bioeconomic evaluation of commercial scale stock enhancement in abalone). Both Department of Fisheries Western Australia projects are funded by the Australian Seafood CRC, collaborate extensively with

this project and meet the CRC's Future Harvest Theme Outcomes 1 (Fisheries management delivering maximum benefit from the resource while maintaining stocks above sustainability indicators) and 2 (Novel management strategies in place which increase economic yield from our fisheries). The projects also targets the CRC's milestone 1.2.2 (Production interventions implemented in at least one fishery) and 1.2.3 (Annual production characterised and interventions optimized in at least one fishery). The projects have been discussed and approved by a range of stakeholders including WAAIA (Western Australia Abalone Industry Association), WAFIC (Western Australian Fishing Industry Council), and ACA (Abalone Council of Australia).

This project has employed a post-doctoral research scientist (PDRS, Jonathan Sandoval-Castillo) with special skills in molecular and population genetics and genomic analysis. The funding for the PDRS and operating for the application of GBS was sourced from Flinders University's and the Department of Fisheries Western Australia funding contribution to the Seafood CRC and the project aims to enhance the two existing Seafood CRC restocking and stock enhancement projects, develop a new skill base at Flinders in the application of GBS to fishery management and aquaculture problems and open opportunities for further projects utilizing GBS for fishery management and gene marker discovery.

1.1 Need

Several initiatives by the Australian Seafood CRC's Future Harvest theme involve some form of stocking or enhancement of fisheries. In WA, populations of Roe's Abalone are currently being restocked after the occurrence of a catastrophic mortality event, while stock enhancement of Greenlip Abalone is also occurring. Translocation of undersized Southern Rocklobster from deep to shallow water locations and ranching of Sea Cucumbers is also occurring. In all such cases there is a need to understand the population genetic dynamics of stocked populations and the extent of genetic interactions of stocked with wild populations. Important considerations are the genetic health of individuals used as breeders, genetic structuring of source populations for restocking, effective population size (i.e. effective number of breeders), effects on inbreeding or loss of genetic diversity, recruitment and geographic spread. Molecular markers can be used to investigate these effects and guide the stocking. Genomic technologies such as GBS are rapidly developing and becoming less expensive and more useful to apply than traditional genetic approaches. These new genomic techniques that directly genotype individuals using sequence information are extremely promising as they eliminate the need for species specific marker development and could provide an efficient and comprehensive means of studying genomes at an individual and population level. Importantly, contrary to traditional genetic methods such as microsatellite surveys, new genomic techniques such as GBS might provide ways of disclosing functional genetic variation (variation that is ecologically relevant and related to the fitness of the individual) or facilitating genomic selection. Once GBS data is mapped to areas of the genome with known function, marker assisted selection for traits of importance to aquaculture becomes possible. Here we engage a post-doc that will specialise in this area and

assist similar projects with Flinders involvement on Roe's and Greenlip Abalone in Western Australia.

1.2 Objectives

- To develop new expertise and transfer genomic methods (i.e. Genotype by Sequencing, GBS) that are currently mostly used for commercially important species of the northern Hemisphere, to benefit the fisheries or aquaculture of key commercial species in Australia;
- To generate genomic data for Australian abalones using a GBS approach known as RAD-seq to analyse changes and patterns in neutral and functionally relevant genetic diversity (i.e. variation related to fitness) associated with restocking and ranching and advise on management strategies for the species concerned;
- Through objective 2, to assist in creating genetically healthy/sustainable restocked populations;
- Through objective 2, to assist in improving returns from ranching while having minimal impact on natural populations;
- Through objective 2, to assist in maintaining healthy fisheries;
- To develop expertise in genetic management associated with enhancement and related interventions.

2. Methods

2.1 Laboratory protocols

2.1.1 Trial of Library construction protocols

Samples from five different tissues of Roe's abalone were used to determine which was the best tissue to use for DNA extraction. Gills and tentacles yielded high quality DNA and therefore it was decided that these tissues would be used for future work. Several RAD library construction protocols were trialled in the laboratory as follows. From this trial run we determined which was the best library construction procedure to use and how many samples could be mixed into each pool for sequencing.

Genomic DNA was extracted from 16 Roe's abalone samples from the Kalbarri Cliffs (remnant population) and 24 samples from the Perth Metropolitan fishery using a modified salting out method. These samples were trialled for variability using RAD-seq (restriction-site associated DNA sequencing), an approach that simultaneously identifies and types tens of 1000s SNPs (single nucleotide polymorphisms).

We prepared three RAD libraries including the 40 samples (16 samples from Kalbarri and 24 samples from Perth) and 8 replicates (4 samples for each population). Briefly, the DNA of each individual was cut using a restriction enzyme (SbfI) producing a set of sticky-end fragments (Fig 1A). These fragments were ligated to adapters (P1) that contain a sequence that bind to an Illumina cell, and a

short sequence (MID) that will unique identify the individual (Fig 1B). After tagging the fragments, all the individuals were pooled in three libraries (each with 12, 12 and 24 individuals). These libraries were sheared to generate fragments with average length of 500 bases. Sheared fragments were ligated to a second adapter (P2) (Fig 1D) and then PCR amplified using P1 and P2 primers (Fig1 E). Subsequently the three libraries were sent to the Genome Quebec Innovation Centre (http://gqinnovationcenter.com/index.aspx?l=e) to be sequenced in an Illumina platform using three lanes.

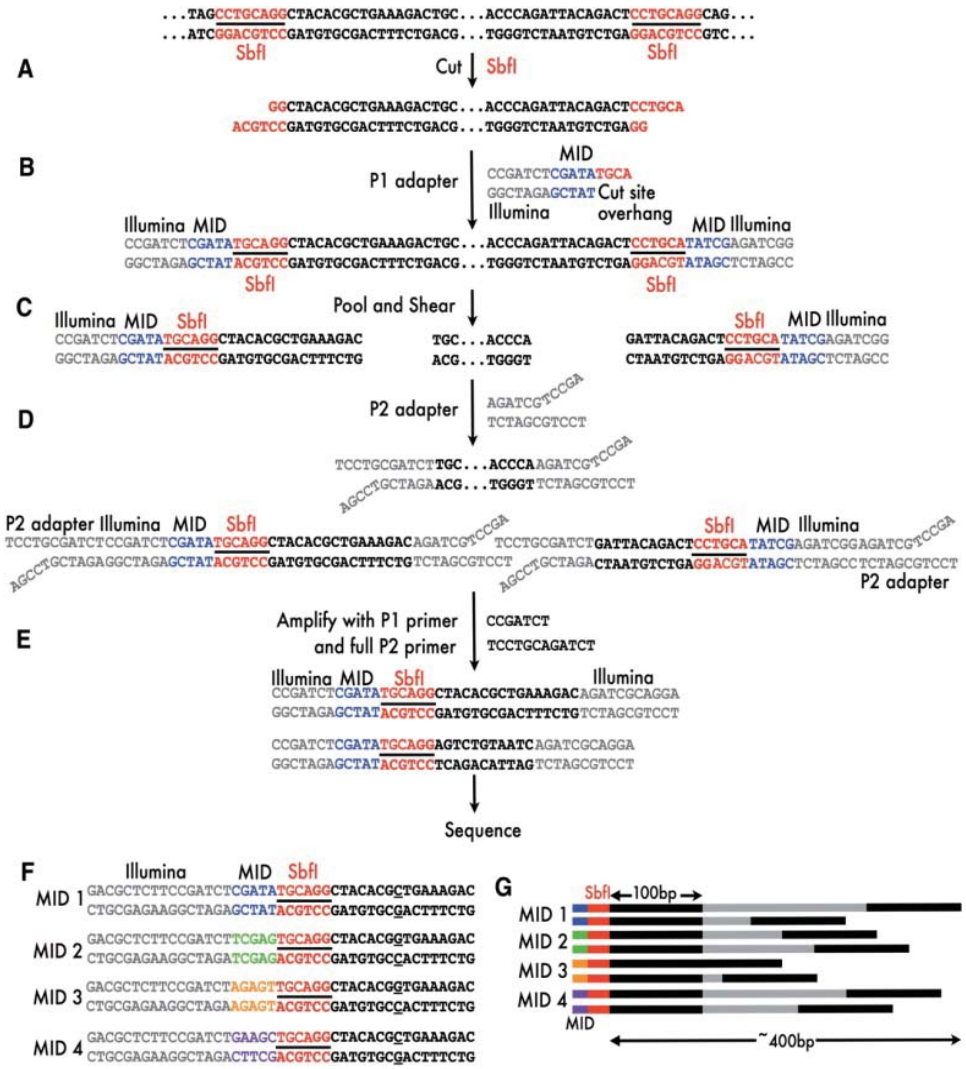


Figure 1: A summary of the RAD-seq process [9].

Given the success of the first RAD-seq trial, a second trial was run in which a total of 48 individual library samples were prepared, pooled and sequenced in a single Illumina lane. This trial was run to determine whether the number of samples to be analysed with RAD-seq could be “ramped-up”.

For Greenlip abalone samples we used a more recently improved version of the orignial RAD protocol known as the double digestion RAD-seq (ddRAD-seq) method [10], which includes a restriction digest with two enzymes simultaneously (SbfI and MseI). This protocol eliminates the random shearing and

introduces a precise selection of genomic fragments by size. This represents a methodological advance compared to the protocol used for the first study species (Roe's abalones). Nonetheless, the results produced by both methods (RAD-seq for Roes and ddRAD-seq for Greenlips) are fully comparable given the context of this project.

2.1.2 Final full-scale library preparations

After the positive results from the trial library constructions, a final set of 816 individual libraries were prepared using the protocol developed at the Molecular Ecology Lab, Flinders University (described above). These libraries consisted of 420 samples of Roe's abalone (*Haliotis roei*) from eight localities (Fig 2); 372 samples of Greenlip abalone (*Haliotis laevigata*) from thirteen localities (Fig 3); and 24 replicates. The libraries were pooled in groups of 48 samples and each of the 17 resulting pools was sequenced in a lane of Illumina HiSeq 2000, a next-generation genome sequencer.

The final large dataset was analysed using the Stacks software pipeline [11] to filter sequences and identify SNPs. Analysis was performed using Flinders University *Colossus*, a cluster of 1,160 CPU cores and 4.25TB (4,250 GB) of RAM – supercomputers are needed to handle the analysis of our large RAD-seq and ddRAD-seq datasets.

2.2 Data analyses

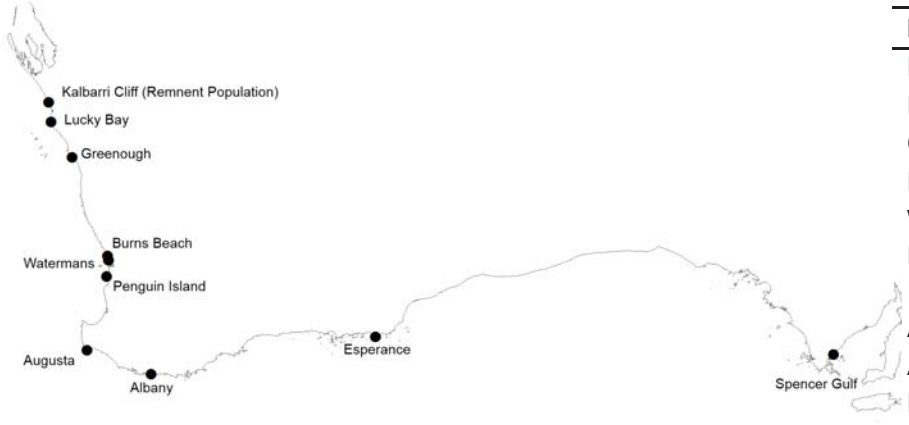
2.2.1 Categorising loci

There are two main forms of genetic diversity, neutral and adaptive variation. Neutral genetic variation is highly valuable for estimation of demographic parameters, particularly connectivity (i.e. gene flow) and population size. By contrast, adaptive (also know as functional) genetic variation affects the organisms ability to adapt to new or changing environments. Therefore, in order to extract the maximum information possible from our genomic data, it is important to be able to discriminate between DNA markers (i.e. loci) that are under selection from those that are neutral loci. We assessed the contribution of natural selection to the overall pattern of genetic differentiation between abalone populations using a F_{ST} outlier approach implemented in ARLEQUIN [12]. Briefly, this method models the expected distribution of the relationship between F_{ST} (Wright's fixation index) and *He* (expected heterozygosity) under an island model of migration with neutral markers. The expected distribution was compared to the observed distribution to identify outlier loci that have excessively high F_{ST}. Such outlier loci are considered likely to be subject to the forces of natural selection.

2.2.2 Genomic analysis

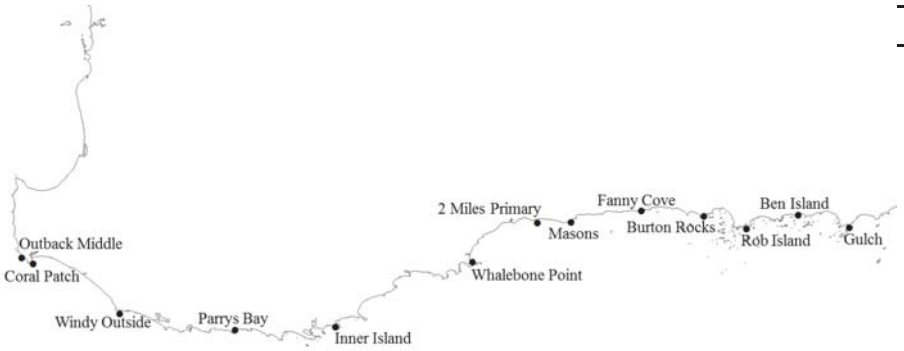
The genetic diversity within localities and the genetic differentiation between localities were estimated using the software ARLEQUIN 3.5 [12]. Software STRUCTURE [13] and ADEGENET [14] were used to determine the optimal number of populations based on our genetic data. STRUCTURE implements a Bayesian clustering algorithm, whereas ADEGENET uses Discriminant Analysis of Principal Components. We tested whether significant genetic differentiation detected between localities could be due to isolation by distance using a Mantel test implemented in GENODIVE 2 [15], which assesses the correlation between a geographic and a genetic distance matrix. To determine the potential

influence of hierarchical population structure, we implemented a stratified Mantel test, in which samples were permutated within each of the three clusters detected by STRUCTURE and ADEGENET.



Locality		n
Kalbarri Cliff	(KC)	50
Lucky Bay	(LB)	50
Greenough	(GN)	50
Burns Beach	(BEA)	50
Watermans	(WM)	50
Penguin Island	(PI)	50
Augusta	(CA)	50
Albany	(AL)	50
Esperance	(ES)	50
Spencer Gulf	(SG)	30

Fig 2. Roe's abalone sampling localities and sample size per locality. These localities cover most of the distribution range of Roe's abalone.



Locality		n
Outback Middle	(OM)	29
Coral Patch	(CP)	28
Windy Outside	(WO)	29
Parrys Bay	(PB)	29
Inner Island	(II)	29
Whalebone Point	(WP)	29
2 Mile Primary	(MP)	28
Masons	(MS)	29
Fanny Cove	(FC)	28
Burton Rocks	(BR)	29
Rob Island	(RI)	28
Ben Island	(BI)	29
Gulch	(GL)	28

Fig 3. Greenlip abalone sampling localities and sample size per locality. These localities cover eight sub-areas in the two main management areas for the commercial Greenlip abalone fishery from Western Australia ([16]).

2.2.3 Seascape analysis

Data for four oceanographic variables (sea surface temperature, oxygen concentration, pH, and nutrient concentration) for the last 100 years were obtained from the NOAA World Ocean Data Base Website (http://www.nodc.noaa.gov/OC5/SELECT/dbsearch/dbsearch.html). For each variable an annual average gridded map at 0.1 degrees resolution was generated using the DIVA algorithm in ODV 4 [17]. To explore the effect of extreme temperatures in the genetic structure, we also generated gridded maps for the average of the maximum annual temperature, and the maximum temperature during the 2010/2011 marine heatwave event. To illustrate environmental variation between sampling sites we performed a principle component analysis (PCA) with the R package FACTOMINER 1.25 [18].

To explore the association between the oceanographic variables and the adaptive genetic differentiation of Roe's abalone populations ("outliers" data set), we applied two multivariable analytical approaches. First, we used the R software ECODIST 1.2.9 [19] to perform a Multiple Regression on Distance Matrices (MRDM) analysis. This is an extension of the partial Mantel test that investigates the relationship between a response distance matrix and any number of explanatory distance matrices. In this case we used the linearized pairwise F_{ST} (F_{ST}/1-F_{ST}) matrix as the dependent variable and the ecological distance matrices as the independent variables. Second, we used a Canonical Correspondence Analysis (CCA) implemented in the R program VEGAN 2.10 [20]. Via constrained ordination diagrams the CCA extracts major synthetic gradients from the response variables in terms of the explanatory variables. In this work we used locality allele frequencies as response variables and the localities specific oceanographic attributes as explanatory variables. Also, a partial CCA was performed using the coordinates of sampling localities as covariates to account for geographic effect in the final model.

3. Results & Discussion

3.1.1 First trial run of RAD-seq technique

We obtained ~550 million reads (small sequences of 100 base pairs long) from our Illumina test run. Approximately 270 million SbfI RADtags were recognized, of which ~22 million were unique sequences. A total of 32,207 SNPs were identified as bi-allelic, with a coverage depth of over 5x (the number of times that a particular locus was sequenced in the same individual) in at least 8 individuals. From the results of the first trial it was apparent that we could genotype sufficient SNP loci using the technique to allow us to perform a powerful seascape genomics analysis to identify ecologically significant adaptive and neutral genomic variation to address the research questions of the *Haliotis roei* restocking program .

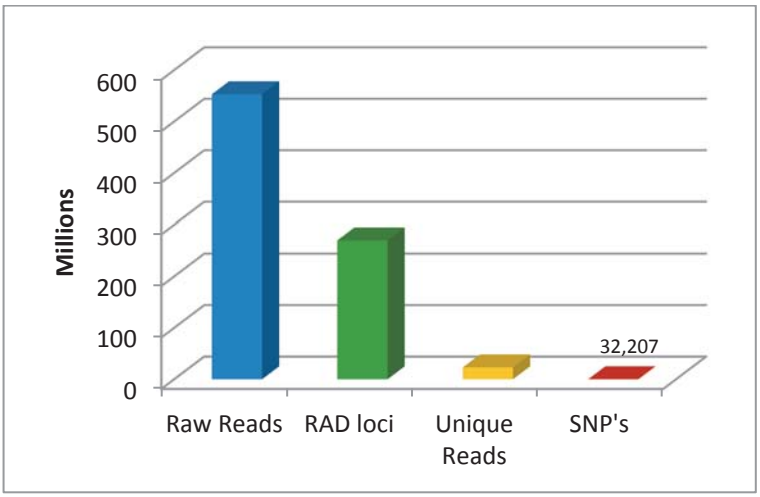


Figure 4: Preliminary results from first *Haliotis roei* RADseq run. Total number of reads, putative RAD loci, unique sequence and SNP's detected

3.1.2 Second test run

From our second test run, we obtained ~145 million sequence reads (small sequences of 100 base pairs long). We used the Stacks software pipeline [11] to filter sequences and identify SNPs, then compared these with the SNP catalogue created from our first test run. Approximately 70 million SbfI RADtags were recognised, of which ~7 million were unique sequences. A total of 6,852 SNPs were bi-allelic, identified in the existing catalogue and had a coverage depth of over 4x in at least 40 of the sequenced individuals. These positive results from the second test run indicated that we can pool 48 samples in one Illumina lane. This allows a greater number of samples to be genotyped with RAD-seq, with minimal increase in cost but also minimal decrease in number of SNP markers.

3.1.3 Full RAD-seq run for Roe's abalones

From the nine Illumina lanes ran using Roe's abalone samples, we obtained ~1.6 billion DNA sequence reads. Each read is a sequence of approximately 100 base pairs long (i.e. we generated a total of ~ 160 billion base pairs of DNA data). After filtering the reads, approximately 774 million SbfI RADtags were recognised, of which ~720 thousand were unique sequences.

3.1.4 Full ddRAD-seq run for Greenlip abalones

From the eight paired-end Illumina lanes ran for the Greenlip abalone samples, we obtained ~3 billion DNA sequence reads (i.e. we generated a total of ~ 300 billion base pairs of DNA data). After filtering the reads, over one billion SbfI RADtags were recognised, of which approximately one million were unique sequences.

3.1.5 Summary of marker variation for the species

For the *H. roei* samples we obtained a total of 31,008 SNP makers (Single Nucleotide Polymophisms); from which we used 9,338 SNP for further analyses because they were bi-allelic, identified in the

existing catalogue, had a coverage depth of over 4x in at least 80% of the sequenced individuals, and were present in all the sampled localities.

For the *H. laevigata* samples we obtained a total of 69,720 SNPs, from which we selected 18,803 SNPs. These SNPS were bi-allelic and had a coverage depth of over 4x in at least 80% of the sequenced individuals. These results strongly indicate that the genomic protocol implemented in this project can be used efficiently for other species of abalone.

3.2 Roe's abalone Results

3.2.1 Genetic diversity

Analysis of the entire SNP data using ARLEQUIN 3.5 [12] showed that levels of genetic diversity were very similar across localities, with slightly higher values near the Perth metropolitan area (Table 1). The diversity detected in Roe's abalone is also comparable with the diversity detected in Greenlip abalone (see Table 6) and the diversity reported in green abalone from California, USA (polymorphic loci =81.5%; expected heterozygosity = 0.27) [21]. Despite the catastrophic mortality associated with the heat wave in the northern stock, there is no evidence of low genetic diversity or population reduction at any sampled locality, including Kalbarri Cliff. Genetic data have low power for detecting a population bottleneck in the first generation after the population reduction. However, due to the very small census population sizes of the remnant population, loss of genetic variation would be detected in the next few generations unless effective re-stocking programs are implemented.

Table 1. Levels of genetic diversity for Roe's abalone from the ten sampled localities. π=nucleotide diversity, He = expected heterozygosity, PL = percentage of polymorphic loci.

		π	He	% PL
Kalbarri Cliff	**KC**	0.15	0.21	85.53
Lucky Bay	**LB**	0.14	0.21	84.62
Greenough	**GN**	0.12	0.22	82.33
Burns Beach	**BEA**	0.16	0.21	87.95
Watermans	**WM**	0.16	0.26	71.40
Penguin Island	**PI**	0.15	0.21	86.38
Augusta	**A**	0.14	0.21	84.51
Albany	**AL**	0.13	0.22	81.05
Esperance	**ES**	0.12	0.22	81.11
Spencer Gulf	**SG**	0.15	0.23	82.62

3.2.2 Categorising loci

We detected 553 outlier loci representing ~5.9% of the scanned loci. Subsequent analyses were conducted for the entire dataset (9,338 SNPs), the "outliers" dataset (553 SNPs) and the "neutral" dataset (8,785 SNPs).

3.2.3 Genetic differentiation

Levels of genetic differentiation were low, but significant (F_{ST} > 0, P<0.01) between most pairs of localities in all the datasets, with highest values in the "outlier" data set (Table 3, 4, 5; Fig4) as tested using ARLEQUIN. Both the STRUCTURE and ADEGENET methods support the existence of one single population for the entire SNP dataset and the "neutral" markers dataset (Fig 5A,B, 6A,B). On the other hand, the "outliers" dataset suggested the existence of at least three markedly differentiated population clusters (Fig 4C, 5C,D, 6C): 1.- A cluster in the northern part of *H. roei* distribution (Kalbarri Cliff and Lucky Bay); 2.- A cluster in the southwest coast of WA (from Greennough to Augusta); and 3.- A cluster in the southern coast of Australia (from Albany to Spencer Gulf) (Fig 7, 8).

Table 3. Levels of genetic differentiation between samples of Roe's abalone from ten localities based on 9338 SNPs. F_{ST} values in bold are significant (P<0.001).

	KC	**LB**	**GN**	**BEA**	**WM**	**PI**	**A**	**AL**	**ES**	**SG**
KC	0.000									
LB	0.001	0.000								
GN	0.002	**0.010**	0.000							
BEA	**0.016**	**0.011**	0.000	0.000						
WM	**0.014**	**0.013**	0.000	0.004	0.000					
PI	**0.014**	**0.012**	0.005	0.004	0.005	0.000				
A	**0.013**	**0.016**	**0.011**	0.004	**0.006**	**0.007**	0.000			
AL	**0.017**	**0.020**	**0.016**	**0.006**	**0.006**	**0.009**	**0.011**	0.000		
ES	**0.009**	**0.016**	**0.015**	0.000	0.000	0.004	**0.009**	0.005	0.000	
SG	**0.011**	**0.013**	**0.007**	**0.008**	**0.007**	**0.008**	**0.007**	0.005	0.000	0.000

Table 4 Levels of genetic differentiation between samples of Roe's abalone from ten localities based on 8785 "neutral" SNPs. F_{ST} values in bold are significant (P<0.001).

	KC	LB	GN	BEA	WM	PI	A	AL	ES	SG
KC	0.000									
LB	0.000	0.000								
GN	0.000	0.009	0.000							
BEA	**0.012**	**0.007**	0.000	0.000						
WM	**0.011**	**0.010**	0.000	0.003	0.000					
PI	**0.011**	**0.010**	0.004	0.003	0.004	0.000				
A	**0.012**	**0.015**	**0.014**	0.004	**0.005**	**0.007**	0.000			
AL	0.005	**0.012**	**0.013**	0.000	0.000	0.002	0.004	0.000		
ES	**0.010**	**0.013**	**0.009**	0.003	**0.005**	**0.006**	**0.010**	**0.008**	0.000	
SG	**0.007**	**0.009**	0.005	**0.006**	0.005	**0.006**	0.005	0.000	**0.005**	0.000

Table 5 Levels of genetic differentiation between samples of Roe's abalone from ten localities based on 553 "outlier" SNPs. F_{ST} values in bold are significant (P<0.001).

	KC	LB	GN	BEA	WM	PI	A	AL	ES	SG
KC	0.000									
LB	**0.018**	0.000								
GN	**0.038**	**0.034**	0.000							
BEA	**0.068**	**0.063**	**0.014**	0.000						
WM	**0.066**	**0.059**	**0.017**	**0.021**	0.000					
PI	**0.058**	**0.050**	**0.018**	**0.021**	**0.013**	0.000				
A	**0.062**	**0.064**	**0.032**	**0.023**	**0.018**	**0.014**	0.000			
AL	**0.084**	**0.078**	**0.045**	**0.037**	**0.022**	**0.027**	**0.027**	0.000		
ES	**0.075**	**0.067**	**0.045**	**0.040**	**0.027**	**0.033**	**0.036**	**0.020**	0.000	
SG	**0.069**	**0.061**	**0.036**	**0.045**	**0.033**	**0.034**	**0.033**	**0.025**	**0.020**	0.000

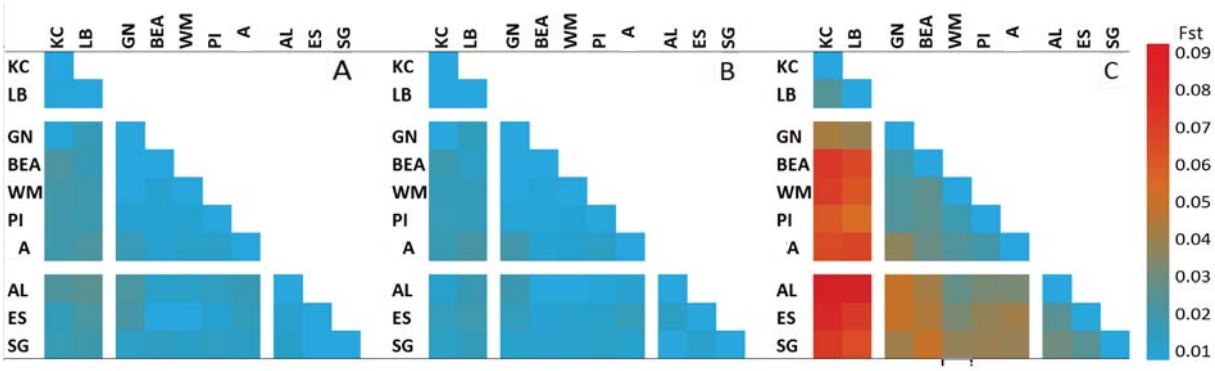


Fig 5. Matrix of pairwise genetic differentiation (F_{ST}). (A) Results based on 9338 SNPs (entire dataset), (B) Results based on 8785 "neutral" SNPs, (C) Result based on 553 outlier SNPs.

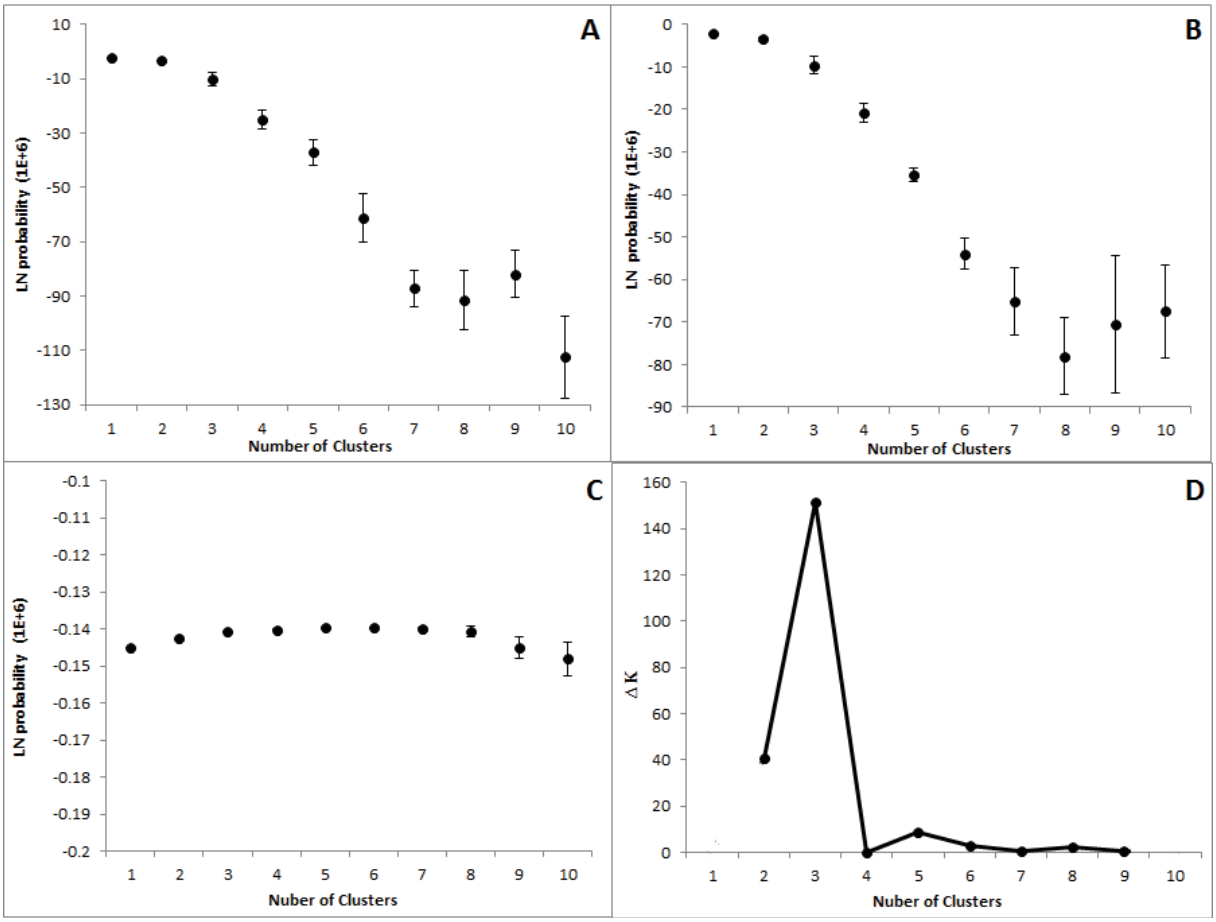


Fig 6. STRUCTURE probability of the data as a function of the number of population clusters (A,B,C); and magnitude of ΔK as a function of number of clusters (D). Results are for three data sets: (A) All the 9338 SNPs; (B) 7875 "neutral" SNPs; (C,D) 553 "outliers". When the highest probability is difficult to define (as in the outlier data set), the highest ΔK (D) should correspond to the optimal number of clusters.

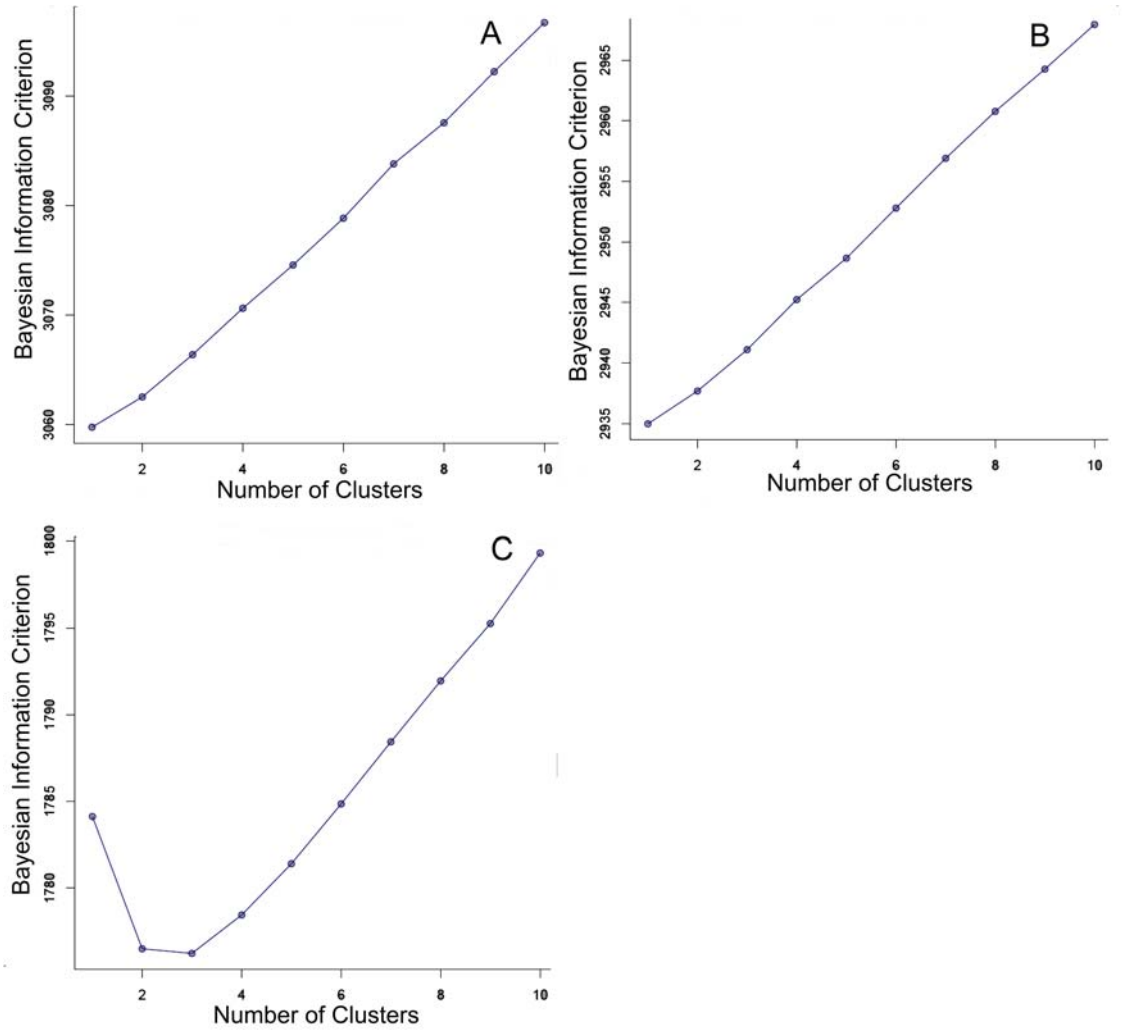


Fig 6 ADEGENET Bayesian Information Criterion as a function of number of clusters: (A) Using all the 9338 SNPs (B) Using 7875 “neutral” SNPs. (C) Using 553 “outliers” SNPs. Ideally, optimal clustering solution should correspond to the lowest Bayesian Information Criterion

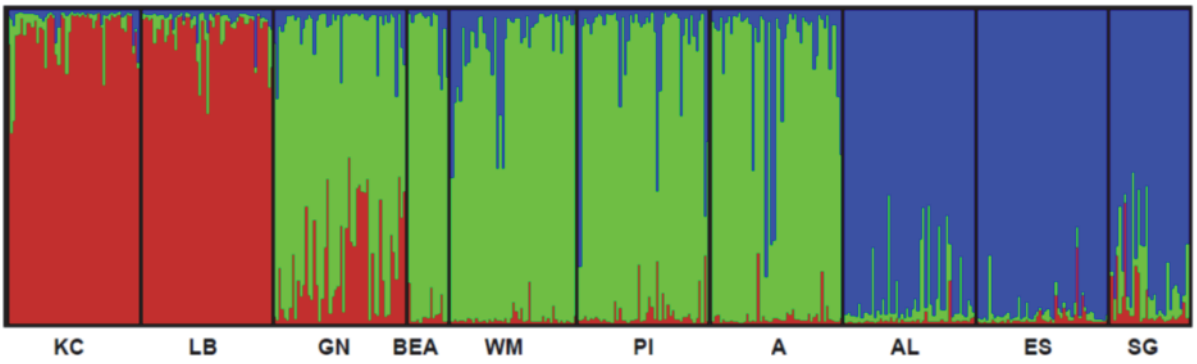


Fig 8 STRUCTURE clustering plot for *Haliotis roei* based on 553 "outlier" SNPs. K=3 is the optimal number of clusters. The figure is based on colour-coded columns where each line corresponds to an individual and the colours to a specific cluster. Black lines separate each sampling locality.

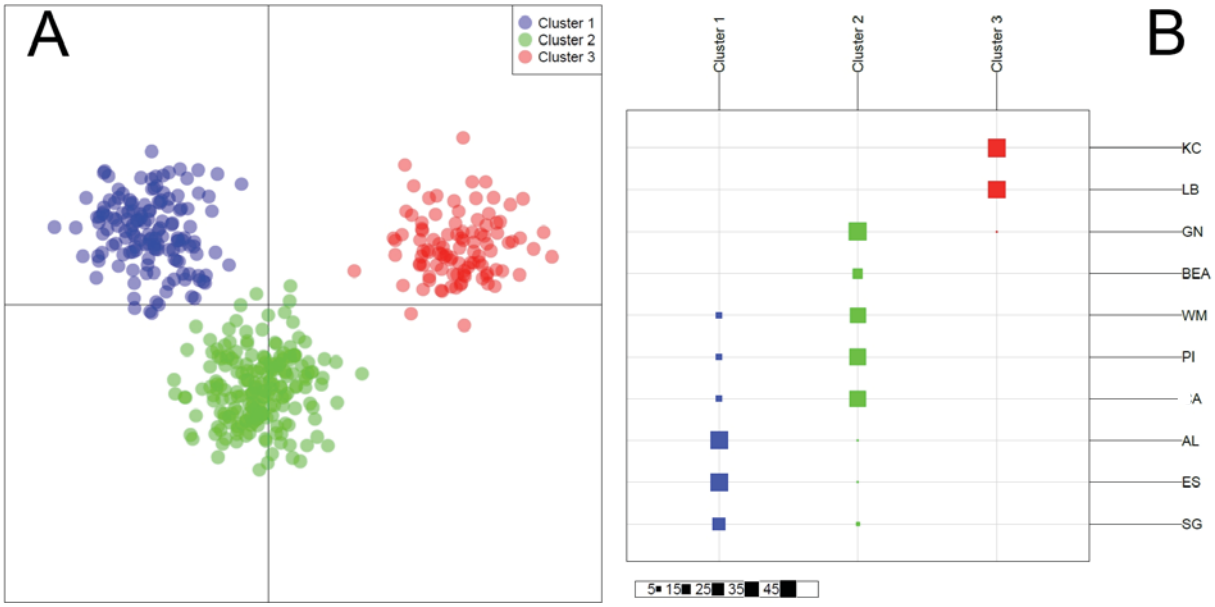


Fig 9 (A) ADEGENET Discriminant analysis of principal components for 553 "outliers" SNPs of *Haliotis roei*. The graphic shows the first two principal components that explain 90% of the genetic variation (PC1=58.8%; PC2=31.2%). (B) Number of samples in each cluster by locality of origin

3.2.4 Isolation by distance

Isolation by distance was only found in WA samples in all datasets (Fig 9).

The stratified Mantel test (in which samples were permuted within each of the three clusters detected by STRUCTURE and ADEGENET) showed that the isolation by distance is only marginally significant for the "neutral" data set (P=0.047) and is not significant for the "outliers" data set (P=0.281).

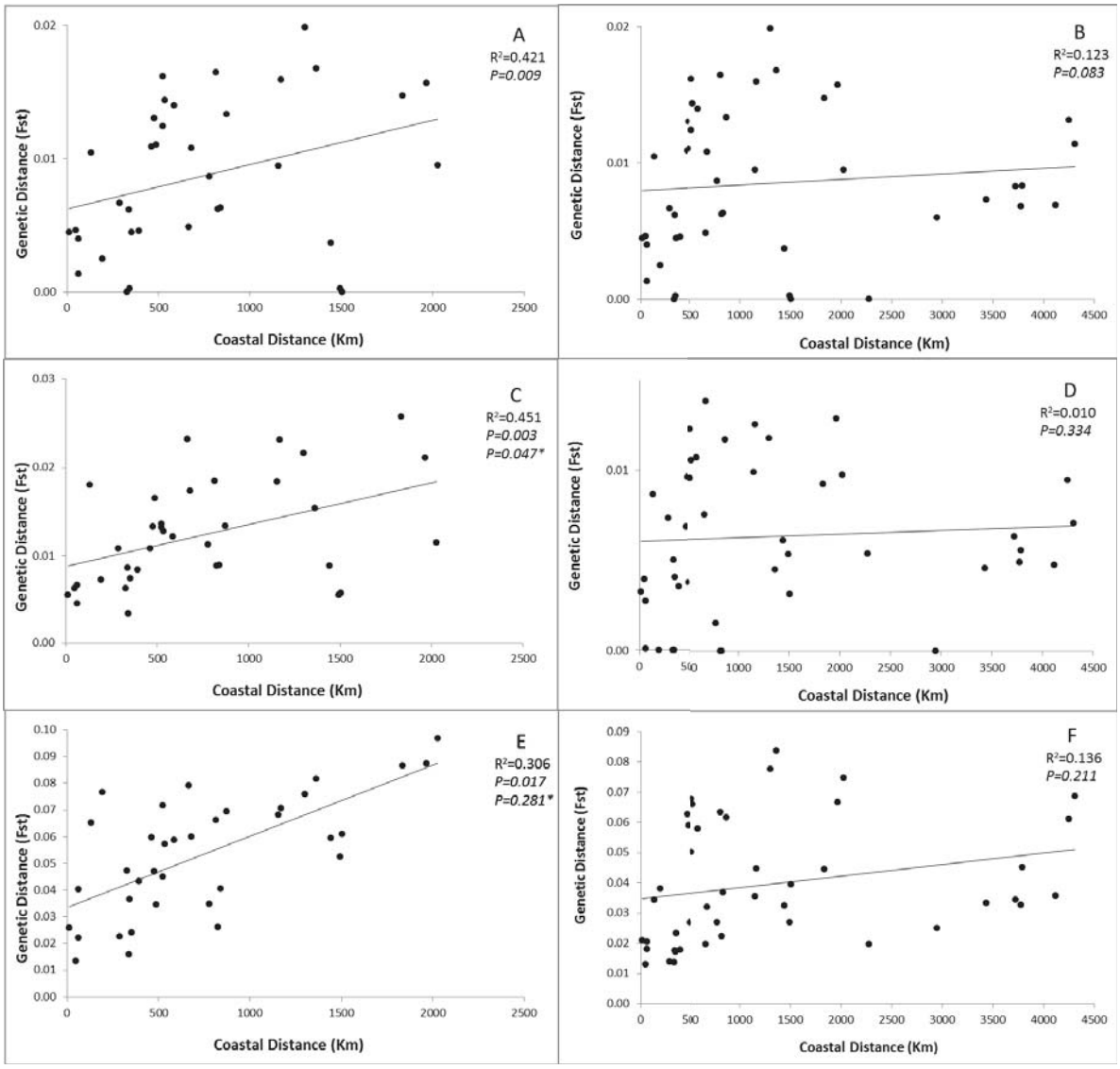


Fig 10 Correlation tests between coastal geographical distance and genetic distance F_{ST} (Mantel test) for pairs of *Haliotis roei* sampling locations. Western Australia samples (A,C,E); including Spencer Gulf samples (B,D,F). Using three data sets: The whole 9338 SNPs data set (A,B); 8785 "neutral" SNPs (C,D); or 553 "outlier" SNPs (E,F). * P values after stratified Mantel test.

3.2.5 Seascape analysis

The PCA of the oceanographic factors revealed three environmentally different regions that are congruent with our three genetic clusters (Fig 10). An exception was Augusta; while this site is genetically clustered with the south west coast of Western Australia, oceanographically it is clustered with south coast of Western Australia.

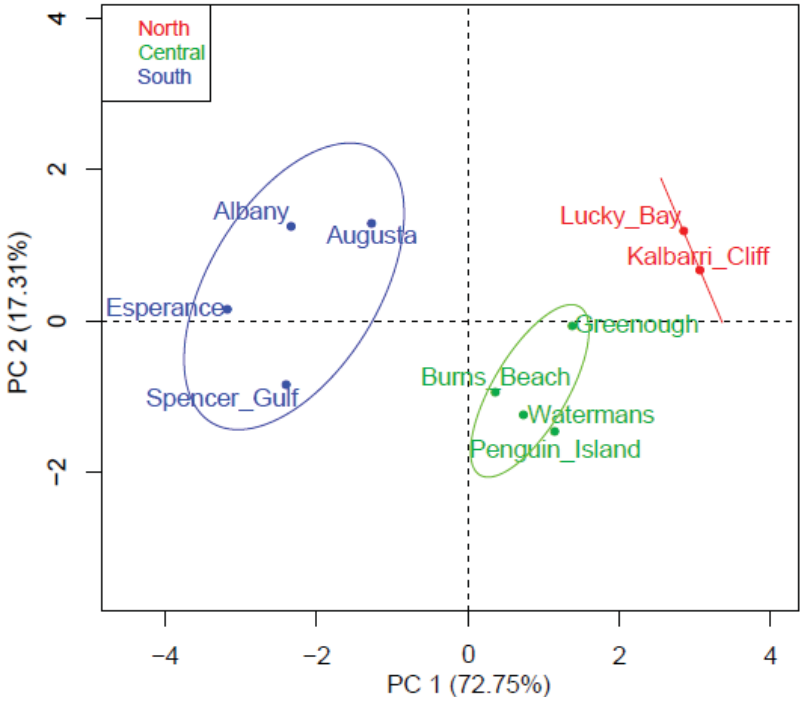


Fig 11 Principal Component Analysis based on six oceanographic variables. The scatterplot shows the first two principal components that explain 90.06% of the variation. Dots are coloured according to the most probable environmental groups. Ellipses represent the 95% confidence level of these groups.

Both the MRDM and the CCA revealed a strong influence of sea surface temperature on the “outlier” genetic pattern (Table 6, 7, Fig 11, 12A). However, after accounting for the spatial effect, only the MRDM was significant (Table 6). The difference between MRDM and CCA results could be due to intrinsic differences between the methods; while MRDM explores the importance of effective separation between the localities, CCA focuses on the relevance of local processes. The results indicate that the individuals are adapted to a temperature range rather than to a specific temperature. Therefore, the difference in temperature between localities is promoting adaptive differentiation and the greater the differences in temperature the greater the genetic differentiation between localities.

Additionally, the lack of significance in the partial CCA can be attributed to the dependence of temperature to latitude rather than to spatial autocorrelation on allele frequency. This hypothesis is supported by the result of the CCA when we include the latitude and longitude as explanatory variables and the temperature as conditional (Table 8). If spatial autocorrelation was the main driver of the allele frequency patterns, both latitude and longitude will be significant even after correcting by temperature.

Table 6. Multiple regressions on distance matrices estimating the correlation of Roe's abalone genetic distance with oceanographic distances. Included are the full model (all oceanographic variables) and a reduced model (oceanographic variables without collinearity). Significant standardised regression coefficients (*b*) after correct for false discovery rate are in bold (*q*<0.05).

Variable	**Full Model**		**Reduced Model**	
	b	*q*	*b*	*q*
Temperature	-0.06	0.82		
Maximum Temperature	**0.66**	**0.01**	**0.84**	**0.01**
Marine Heatwave Temperature	0.58	0.12		
Oxygen Concentration	0.33	0.56	0.24	0.18
Nutrients Concentration	-0.28	0.12	-0.22	0.14
pH	0.09	0.56	0.11	0.36
Geographic Distance	-0.50	0.16	-0.10	0.50
Model	0.84	**0.00**	0.82	**0.00**

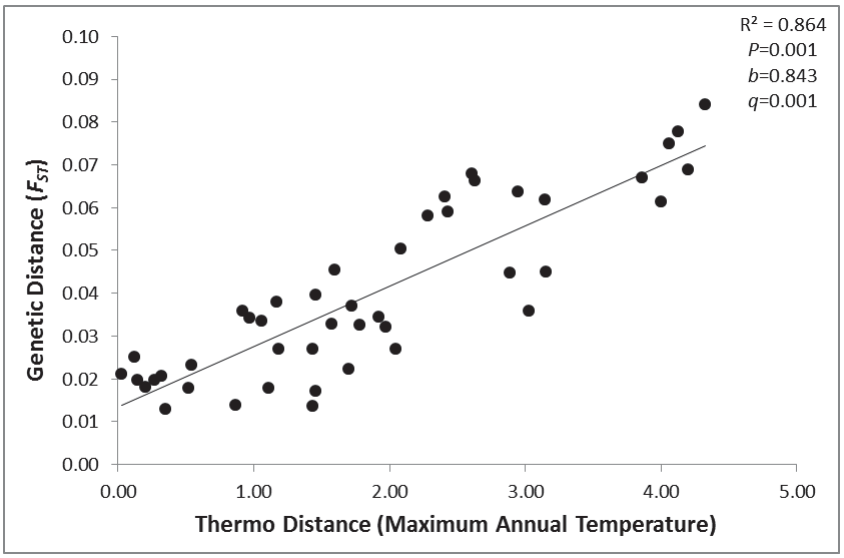


Fig 12 Correlation tests between thermo distance and genetic distance for pairs of *Haliotis roei* sampling locations. Regression coefficient (R^2) and standardized regression coefficient (*b*) with their associated significance *P* and *q* values

Table 7. Canonical Correspondence Analysis exploring the relation between Roe's abalone allele frequencies of 538 outliers SNPs and six oceanographic variables. Simple CCA and partial CCA (geographic coordinates as conditionals). Significant canonical coefficients after correcting for false discovery rate are in bold (q<0.05).

Variables	CCA				Partial CCA			
	CCA1	CCA2	q	Variation explained (%)	CCA1	CCA2	q	Variation explained (%)
Temperature	**-0.73**	**0.50**	**0.01**		0.07	-0.06	0.46	
Maximum Temperature	-0.81	0.55	0.36		-0.09	0.04	0.45	
Marine Heatwave Max Temperature	-0.66	0.70	0.42		0.02	-0.01	0.54	
Oxygen Concentration	0.66	-0.54	0.28		-0.14	0.00	0.24	
Nutrients Concentration	0.21	-0.75	0.07		-0.41	0.22	0.09	
pH	-0.81	0.01	0.36		-0.62	0.08	0.35	
Geographic Connectivity	0.03	-0.58	0.45		-0.06	0.07	0.71	
Model	**0.37**	**0.14**	**0.02**	**86.03**	0.32	0.17	0.38	61.56
Conditionals (Latitude, Longitude)								38.44

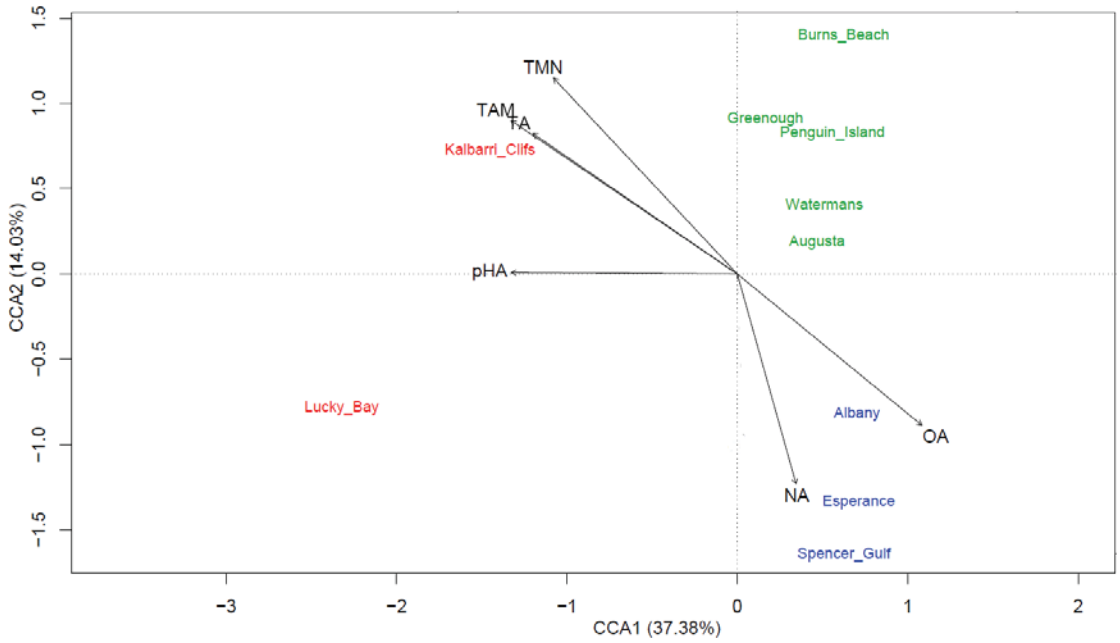


Fig 13 Canonical Correspondence Ordination based on six oceanographic variables and 538 "outlier" SNPs. The scatterplot shows sampling sites in relation to the first two canonical components, which explain 51.42% of the variation.

Table 8. Canonical Correspondence Analysis exploring the relation between Roe's abalone allele frequencies of 538 outliers SNPs and geographic space of sampling sites. Simple CCA and partial CCA (temperature as conditional). Significant canonical coefficients after correcting for false discovery rate are in bold (q<0.05).

Variables	**CCA**				**Partial CCA**			
	CCA1	**CCA2**	**q**	**Variation explained (%)**	**CCA1**	**CCA2**	**q**	**Variation explained (%)**
Latitude	**-1.00**	**0.01**	**0.00**		-0.16	-0.39	0.23	
Longitude	0.51	-0.86	0.21		-0.60	0.71	0.11	
Model	**0.28**	**0.12**	**0.00**	**40.04**	0.17	0.13	0.10	22.80
Conditional (Temperature)								25.70

Based on the results of analyses described above, and on the scenario that the Lucky Bay population is also depleted in numbers, we suggest it would best to use individuals from Greenough to re-establish the Roe's abalone at Kalbarri Cliff and Lucky Bay localities. Greenough shows some genetic admixture with the northern localities and it is part of the most genetically similar cluster.

3.3 Greenlip abalone Results

3.3.1 Genetic diversity

Similarly to what was found for Roe's abalone, the levels of genetic diversity in Greenlip abalone were very similar across sampled localities, with marginally higher values for the eastern most localities (Table 9). The genetic diversity detected in Greenlip abalone was slightly higher than that from Roe's abalone and similar to the diversity reported for green abalone from California, USA [21]. We found no evidence pointing to reductions of genetic diversity, as would be expected if the fishery for this species was overexploited.

Table 9. Levels of genetic diversity for Greenlip abalone in the ten sampled localities. π=nucleotide diversity, He = expected heterozygosity, PL = percentage of polymorphic loci.

		π	He	% PL
Outback Middle	**OM**	0.19	0.26	0.7448
Coral Patch	**CP**	0.19	0.27	0.7505
Windy Outside	**WO**	0.17	0.28	0.7447
Parrys Bay	**PB**	0.20	0.27	0.7739
Inner Island	**II**	0.19	0.27	0.7570
Whalebone Pt	**WP**	0.20	0.27	0.7677
2 Mile Primary	**2MP**	0.20	0.27	0.7682
Masons	**MS**	0.18	0.26	0.7483
Fanny Cove	**FC**	0.23	0.29	0.7851
Burton Rocks	**BR**	0.21	0.27	0.8034
Rob Island	**RI**	0.20	0.26	0.7762
Ben Island	**BI**	0.19	0.25	0.7777
Gulch	**GL**	0.17	0.28	0.7402

3.3.2 Categorising loci

We detected a greater number of outlier loci for Greenlip abalone (1,026 outliers) than for Roe's abalone (553), but the proportion of outliers with respect to the scanned loci was similar for both species (-5.4% Greenlip abalone, ~5.9% Roe's abalone). Subsequent analyses were conducted for the entire dataset (18,803 SNPs), the "outlier" dataset (1026 SNPs) and the "neutral" dataset (17,777 SNPs) of the Greenlip abalone.

3.3.3 Genetic differentiation

Levels of genetic differentiation for Greenlip abalone ranged from low to nil between most localities for both the entire and the neutral datasets. However, for the "outlier" dataset the levels of differentiation ranged from low to high (Tables 10, 11, 12; Fig 13). Both the STRUCTURE and ADEGENET methods supported the existence of a single population based on both the entire SNP and the "neutral" datasets (Fig 14A,B, 15A,B). The "outliers" dataset suggested the existence of at least five differentiated population clusters (Fig 13C, 14C,D, 14C): 1- the western part of the *H. laevigata* distribution (from Outback to Windy Outside); 2- the Albany sub-area (Parrys Bay and Whalebone Port); 3- the Hopetoun sub-area (from Inner Island to Mason); 4 - the West sub-area (Fanny Cove and Burton Rocks); and 5 - the eastern sampling area (from Rob Island to Gulch) (Fig 16, 17).

Table 10. Levels of genetic differentiation between samples of Greenlip abalone from thirteen localities based on 18,803 SNPs. F_{ST} values in bold are significant (P<0.001).

	OM	CP	WO	PB	WP	II	2MP	MS	FC	BR	RI	BI	GL
OM	0.000												
CP	0.000	0.000											
WO	0.001	0.000	0.000										
PB	0.002	0.001	0.000	0.000									
WP	0.007	0.006	0.006	0.000	0.000								
II	**0.016**	**0.016**	**0.013**	0.004	0.003	0.000							
2MP	**0.015**	**0.015**	0.012	0.002	0.003	0.000	0.000						
MS	**0.012**	**0.013**	**0.016**	0.000	0.001	0.000	0.000	0.000					
FC	**0.014**	0.012	0.000	0.004	0.012	**0.017**	**0.014**	0.012	0.000				
BR	0.007	0.004	0.000	0.000	0.004	**0.014**	0.009	0.007	0.000	0.000			
RI	0.007	0.007	0.004	0.000	0.003	0.010	0.007	0.006	0.001	0.000	0.000		
BI	0.007	0.007	0.009	0.001	0.004	0.011	0.008	0.007	0.004	0.001	0.000	0.000	
GL	0.007	0.005	0.009	0.000	0.001	0.009	0.006	0.010	0.000	0.000	0.000	0.000	0.000

Table 11 Levels of genetic differentiation between samples of Greenlip abalone from thirteen localities based on 17,777 “neutral” SNPs. F_{ST} values in bold are significant (P<0.001).

	OM	CP	WO	PB	WP	II	2MP	MS	FC	BR	RI	BI	GL
OM	0.000												
CP	0.000	0.000											
WO	0.000	0.000	0.000										
PB	0.000	0.000	0.000	0.000									
WP	0.002	0.001	0.000	0.000	0.000								
II	0.003	0.002	0.000	0.000	0.000	0.000							
2MP	0.004	0.003	0.000	0.000	0.000	0.000	0.000						
MS	0.002	0.001	0.004	0.000	0.000	0.000	0.000	0.000					
FC	**0.009**	**0.008**	0.000	0.000	0.004	0.000	0.000	0.000	0.000				
BR	0.003	0.001	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000			
RI	0.004	0.003	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000		
BI	0.003	0.003	0.004	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	
GL	0.002	0.001	0.004	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000

Table 12 Levels of genetic differentiation between samples of Greenlip abalone from thirteen localities based on 1,026 “outlier” SNPs. F_{ST} values in bold are significant (P<0.001).

	OM	CP	WO	PB	WP	II	2MP	MS	FC	BR	RI	BI	GL
OM	0.000												
CP	0.016	0.000											
WO	**0.046**	0.031	0.000										
PB	**0.054**	**0.057**	**0.064**	0.000									
WP	**0.072**	**0.080**	**0.099**	0.011	0.000								
II	**0.185**	**0.203**	**0.233**	**0.117**	**0.082**	0.000							
2MP	**0.157**	**0.169**	**0.189**	**0.082**	**0.060**	0.040	0.000						
MS	**0.149**	**0.165**	**0.186**	**0.074**	**0.054**	0.037	0.012	0.000					
FC	**0.089**	**0.076**	**0.051**	**0.092**	**0.125**	**0.260**	**0.204**	**0.200**	0.000				
BR	**0.068**	**0.053**	0.042	**0.066**	**0.101**	**0.237**	**0.184**	**0.182**	0.012	0.000			
RI	**0.055**	**0.061**	**0.077**	**0.058**	**0.067**	**0.182**	**0.145**	**0.137**	**0.058**	0.045	0.000		
BI	**0.059**	**0.068**	**0.084**	**0.054**	**0.061**	**0.161**	**0.128**	**0.116**	**0.066**	**0.058**	0.021	0.000	
GL	**0.072**	**0.078**	**0.088**	**0.064**	**0.070**	**0.178**	**0.143**	**0.138**	**0.071**	**0.058**	0.029	0.019	0.000

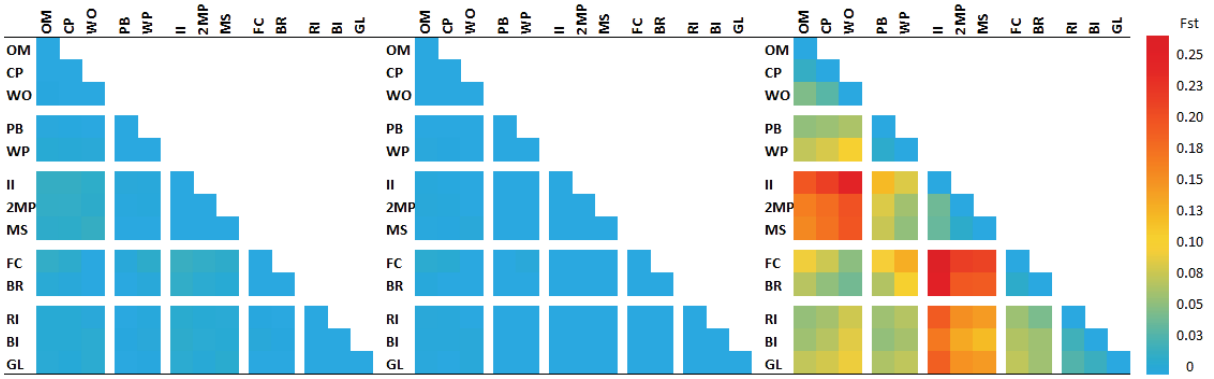


Fig 14. Matrix of pairwise genetic differentiation (F_{ST}). (A) Results based on 18,803 SNPs (entire dataset); (B) Results based on 17,777 “neutral” SNPs; (C) Results based on 1,026 outlier SNPs.

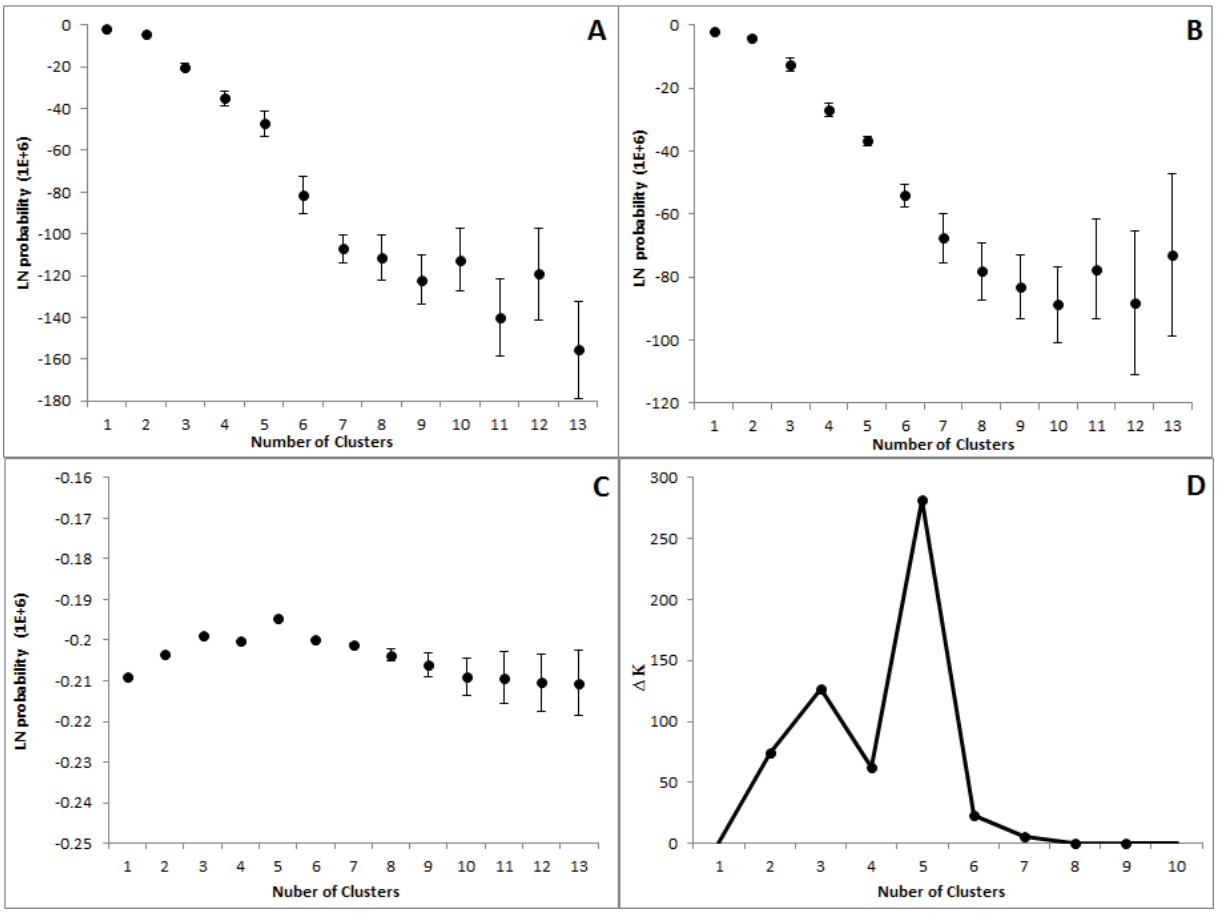


Fig 15. STRUCTURE probability of the data as a function of the number of population clusters (A, B, C); and magnitude of ΔK as a function of number of clusters (D). Results shown for the three data sets: (A) all 18,803 SNPs; (B) 17,777 “neutral” SNPs; (C,D) 1,026 “outliers”. When the highest probability is difficult to define (as in the outlier data set), the highest ΔK (D) should correspond to the optimal number of clusters

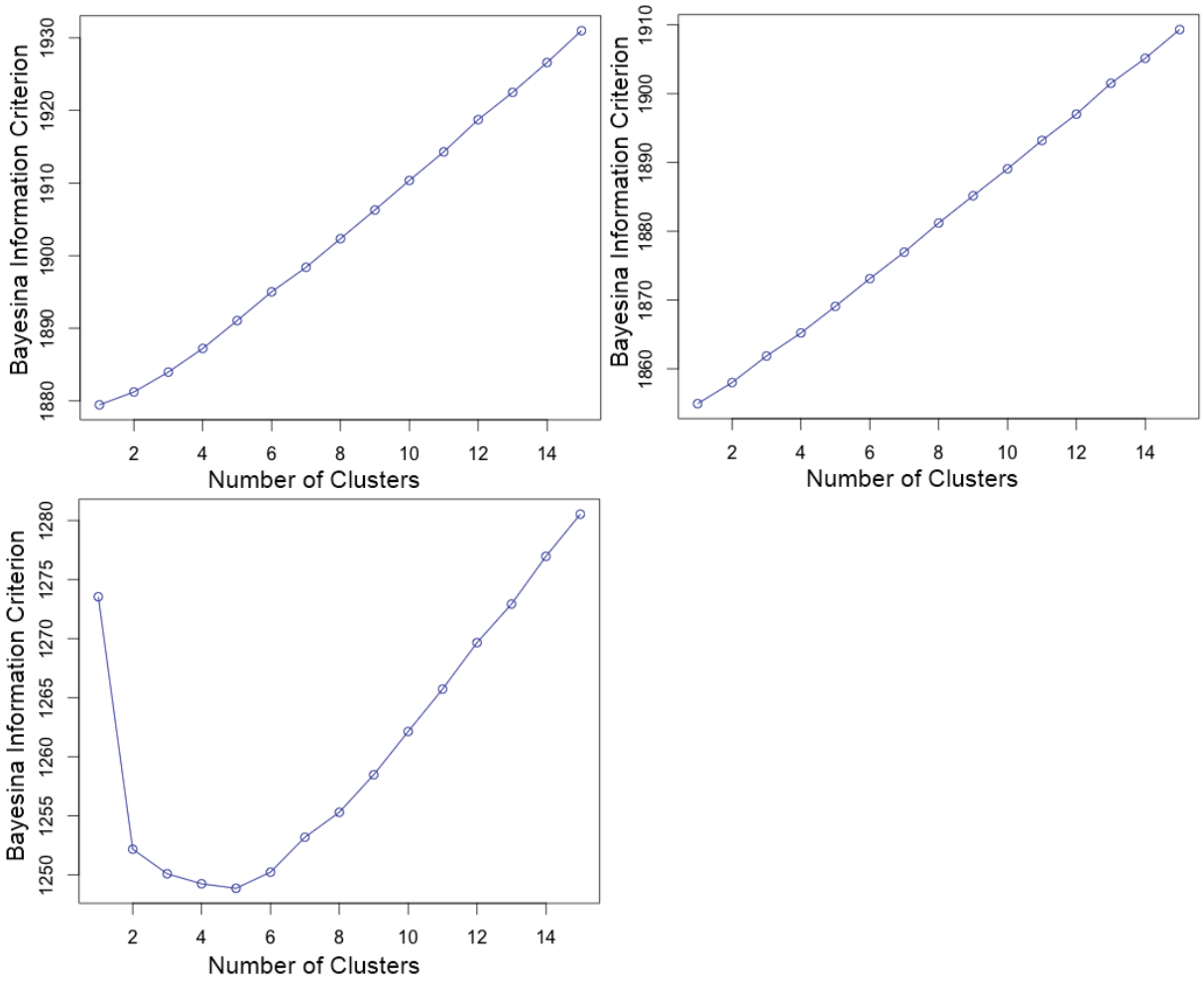


Fig 16 ADEGENET Bayesian Information Criterion as a function of number of clusters: (A) using all the 18,803 SNPs; (B) using 17,777 “neutral” SNPs; (C) using 1,026 “outliers” SNPs. Ideally, optimal clustering solution should correspond to the lowest Bayesian Information Criterion.

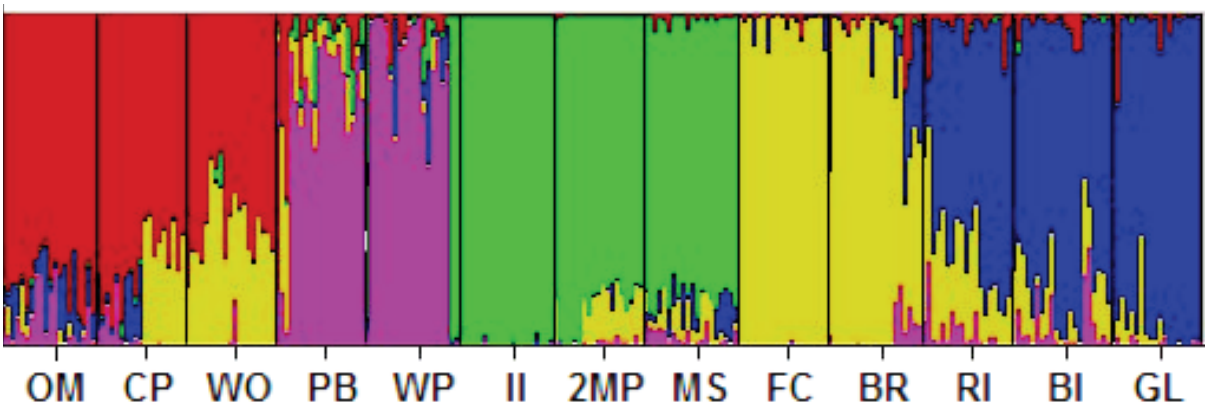


Fig 17 STRUCTURE clustering plot for *Haliotis laevigata* based on 1,026 "outlier" SNPs. K=5 is the optimal number of clusters. The figure is based on colour-coded columns where each line corresponds to an individual and the colours to a specific cluster. Black lines separate each sampling locality

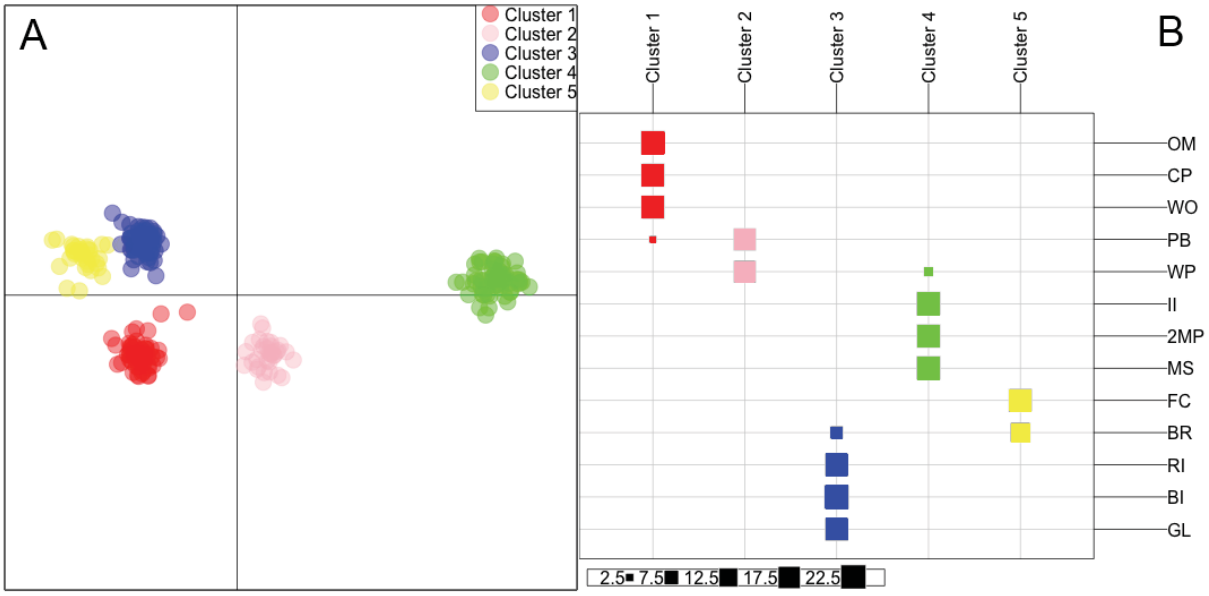


Fig 18 (A) ADEGENET Discriminant analysis of principal components for 1026 "outliers" SNPs of *Haliotis laevigata*. The graphic shows the first two principal components that explain 91.5% of the genetic variation (PC1=83.0%; PC2=8.5%). (B) Number of samples assigned to different clusters by locality of origin.

3.3.4 Isolation by distance

Isolation by distance was only found in the "neutral" dataset (*P*=0.01;Fig 18B).

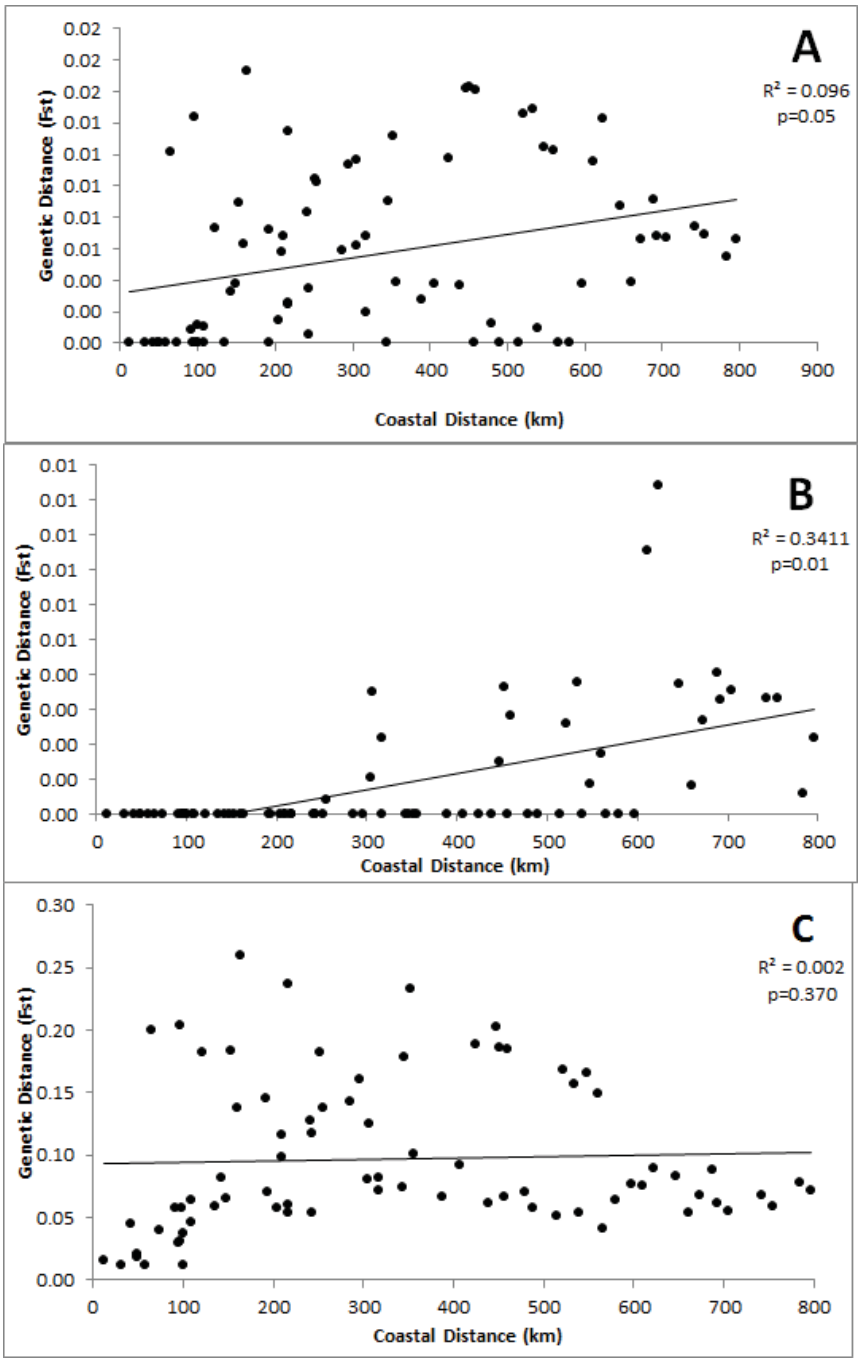


Fig 19 Correlation tests between coastal geographical distance and genetic distance F_{ST} (Mantel test) for pairs of *Haliotis laevigata* sampling locations. 18,803 SNPs data set (A); 17,777 "neutral" SNPs (B); 1,026 "outlier" SNPs (C).

3.3.5 Seascape analysis

The PCA of the annual oceanographic data revealed four environmentally different regions that are partially congruent with our five genetic clusters (Fig 19). Inner Island samples are very different genetically to those from other localities in the Albany region. However, geographically and oceanographically this locality clusters with Parrys Bay and Whalebone Point. In addition, Fanny Cove and Burton Rocks compose an unique cluster but oceanographically they clustered with 2 Mile

Primary and Masons. The overall congruence between oceanographic and genetic clusters indicate a strong influence of environmental factors in the genetic structure of the Greenlip abalone.

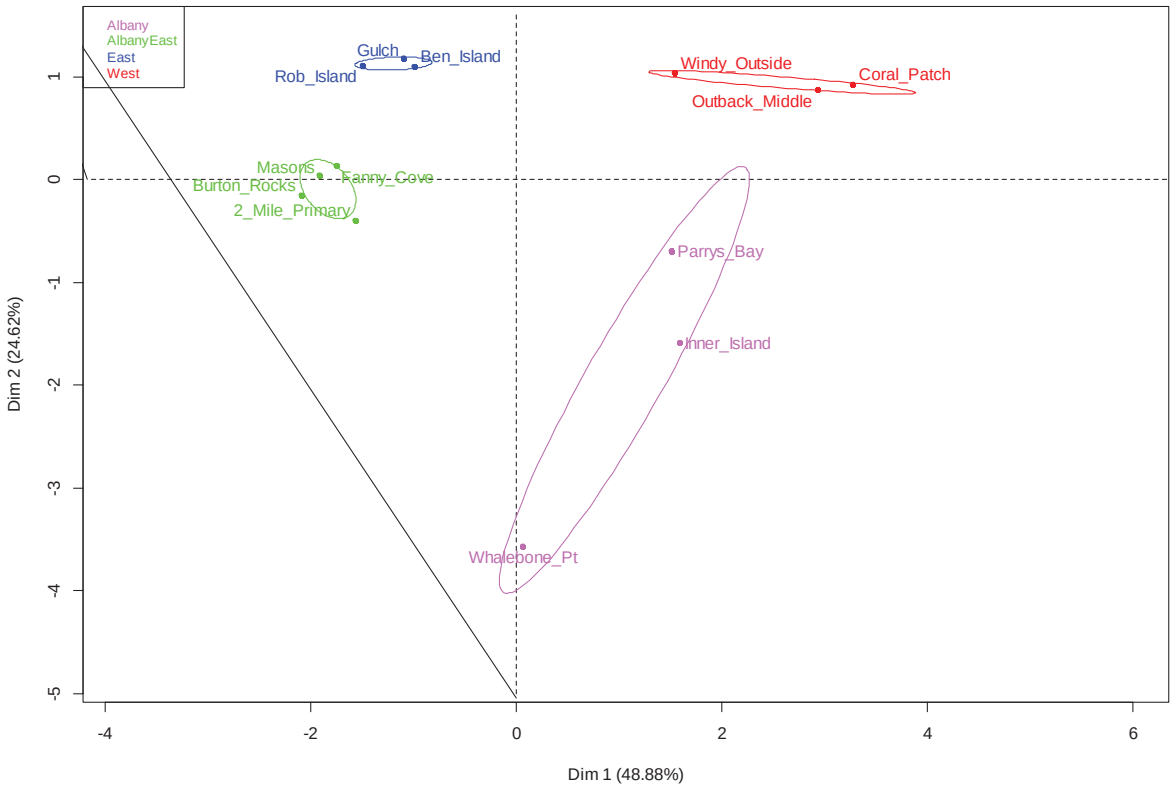


Fig 20 Principal Component Analysis based on five oceanographic variables. The scatterplot shows the first two principal components that explain 73.5% of the variation. Dots are coloured according to the most probable environmental groups. Ellipses represent the 95% confidence level of these groups.

The MRDM analysis shows statistically significant correlations of sea surface temperature and oxygen concentration with the “outlier” genetic pattern (Table 13). However, when we remove collinear variables, only oxygen concentration appears correlated with levels of genetic differentiation. These results indicate that differences in oxygen concentration between localities are promoting adaptive differentiation between these groups of populations.

The CCA did not show significant correlation with any of the oceanographic variables (Table 14), suggesting that, as in Roe’s abalone the adaptive differentiation observed in Greenlip samples is associated with the difference between localities, with individuals adapted to an oxygen concentration range rather than a specific oxygen concentration.

Table 13. Multiple regression on distance matrices estimating the correlation of Greenlip abalone genetic distance with oceanographic distances. Included is a full model (all oceanographic variables) and a reduced model (oceanographic variables without collinearity). Significant standardised regression coefficients (*b*) after correction for false discovery rate are in bold (*q*<0.05).

Variable	Full Model		Reduce Model	
	b	*q*	*b*	*q*
Temperature	**-0.55**	**0.04**	-0.21	0.07
Maximum Temperature	0.11	0.48		
Oxygen Concentration	**0.41**	**0.03**	**0.38**	**0.04**
Nutrients Concentration	0.22	0.18	0.30	0.06
pH	-0.12	0.51	-0.13	0.39
Geographic Distance	0.41	0.22		
Model	**0.23**	**0.04**	**0.22**	**0.03**

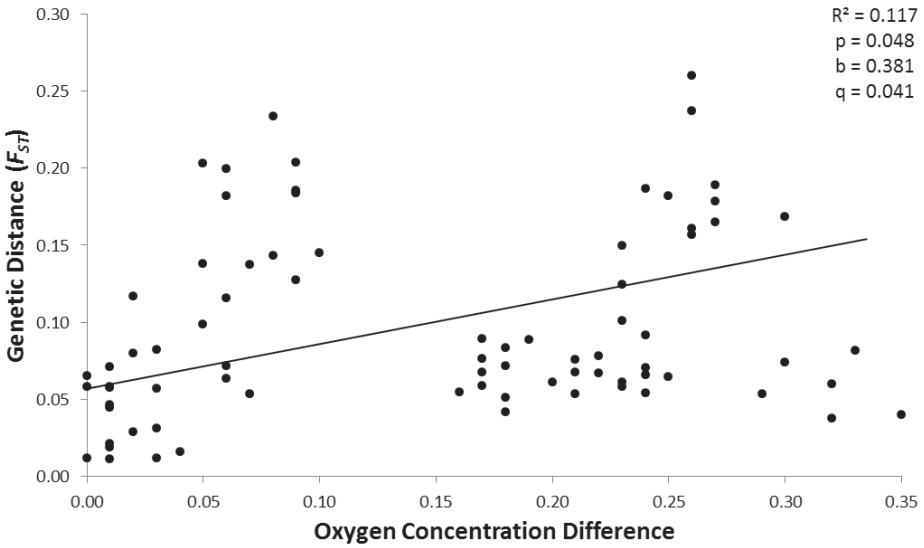


Fig 21 Correlation tests between oxygen difference and genetic distance for pairs of *Haliotis laevigata* sampling locations. Regression coefficient (R^2) and standardised regression coefficient (*b*) with their associated *P* and *q* values.

The weaker correlation between adaptive differentiation and oceanographic factors in Greenlip abalones compared to Roe's abalones could be due to the smaller range of variation in temperature and oxygen concentration observed along the Greenlip's sampling area. While the Roe's sampling area includes differences between localities greater than 3°C in temperature and 1 mg/L in oxygen concentration, the Greenlip's sampling area shows maximum differences of 1.3°C in temperature and 0.4 mg/L in oxygen concentration (See Table 1S).

Table 14. Canonical Correspondence Analysis (CCA) exploring the relationship between Greenlip abalone allele frequencies of 1026 “outlier” SNPs and five oceanographic variables. Shown are results for simple CCA and partial CCA (geographic coordinates as conditionals). Significant canonical coefficients after correcting for false discovery rate are in bold (q<0.05).

Variables	**CCA**				**Partial CCA**			
	CCA1	**CCA2**	***q***	**Variation explained (%)**	**CCA1**	**CCA2**	***q***	**Variation explained (%)**
Temperature	-0.26	0.87	0.75		-0.19	0.25	0.46	
Maximum Temperature	0.096	0.607	0.25		0.16	0.41	0.15	
Oxygen Concentration	0.023	-0.73	0.37		0.03	-0.21	0.25	
Nutrients Concentration	-0.74	0.527	0.11		-0.61	0.35	0.46	
pH	-0.54	-0.79	0.11		-0.46	0.14	0.14	
Model	0.031	0.014	0.06	53.21	**0.42**	**0.11**	**0.046**	**71.3**
Conditionals (Latitude, Longitude)								17.78

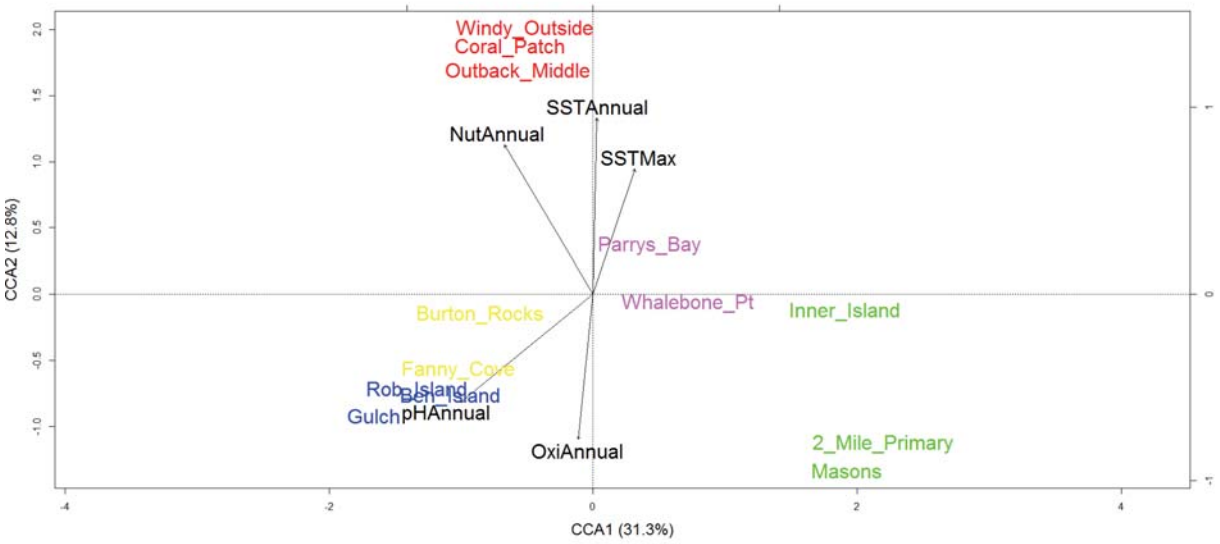


Fig 22 Canonical Correspondence Ordination based on five oceanographic variables and 1026 “outlier” SNPs. The scatterplot shows sampling sites in relation to the first two canonical components, which explain 44.1% of the variation.

These results show some concordance with the fisheries management sub-areas (see Fig 23). Individuals from localities within Augusta and Windy Harbour sub-areas can be used as broodstock for the western populations, while individuals from localities within Town, Duke and Arid sub-areas can be used for stock enhancement in the eastern populations. However, Albany, Hopetoun and West sub-areas form independent adaptive genetic clusters, therefore broodstock should be taken directly from the populations that are to be enhanced.

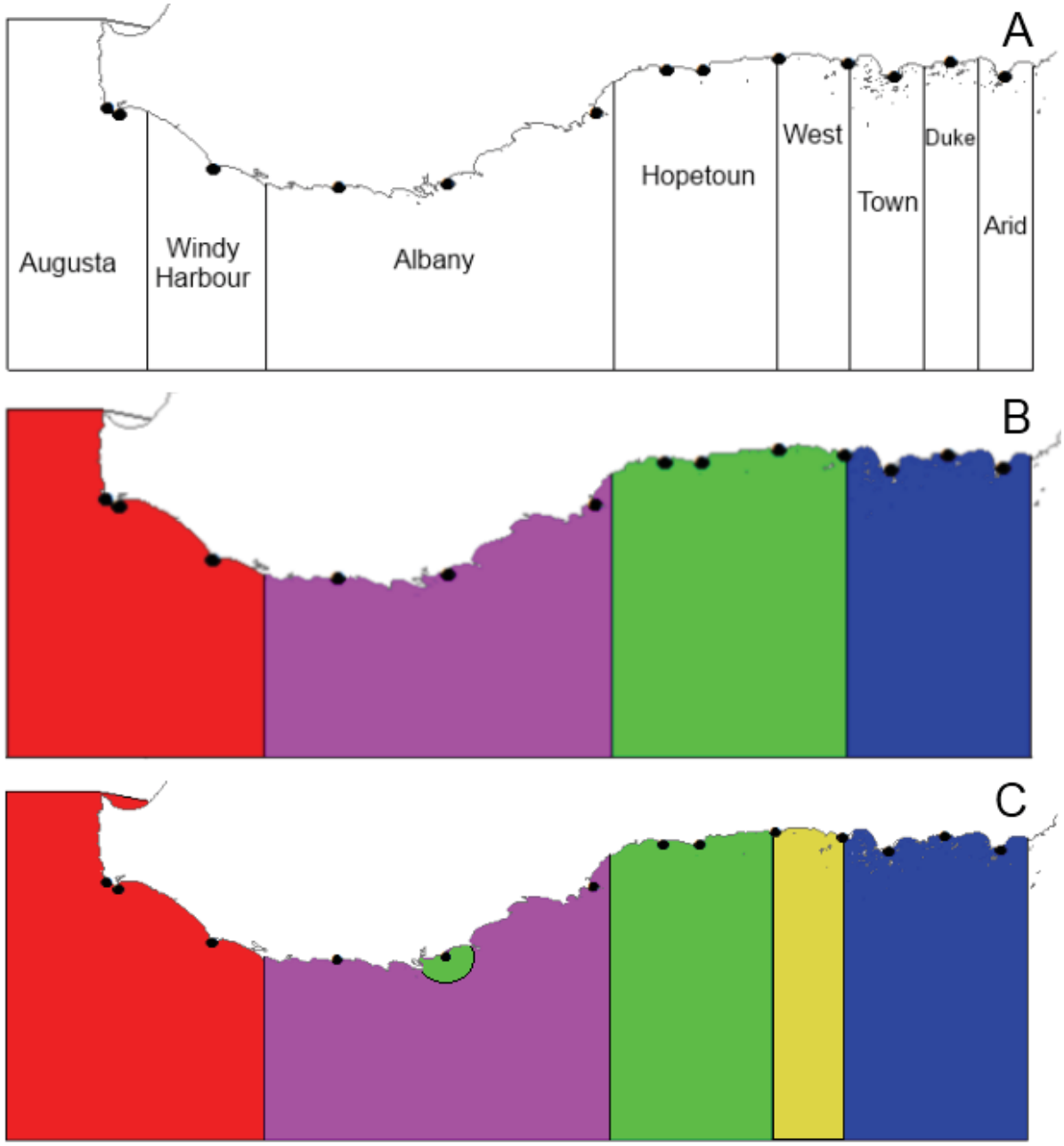


Fig 23 Maps showing A) Management sub-areas for the commercial Greenlip fisheries covered in this report. B) Oceanographic regions detected by our PCA based on six oceanographic variables. C) Geographic distribution of the Greenlip genetic clusters based on 1,026 “outlier” SNPs.

3.4 Results summary

Our results suggest that both Roe's abalone and Greenlip abalone show marked genetic structure as a result of local adaptation. A minimum of three adaptive clusters of Roe's abalone populations and five adaptive clusters of Greenlip abalone should be considered for conservation management of these species. For instance, to re-establish the Roe's abalone at the Kalbarri Cliff locality and Lucky Bay, it would be best to use animals from Greenough, since its cluster shows the least amount of genetic differentiation with the northern cluster and since this locality shows some admixture with the two northern localities. For the Greenlip abalone stock enhancement program, broodstock should be taken from localities occurring within the same adaptive genetic cluster of the stock to be enhanced. The adaptive genetic differentiation found in our samples is correlated with the oceanography of the region. Seascape analyses support a scenario of a gradient in thermal adaptation in Roe's abalone populations, and oxygen related adaptation in Greenlip abalone populations. The patterns of adaptive differentiation found in this study should be examined further using transcriptome-outlier annotation. Some of the outlier genes identified could be affecting the fitness of individuals in these different environments and some of the SNPs identified could be causative variants affecting these traits. The identification of genes associated with temperature or oxygen through this approach will represent an important step to understanding heat-stress or hypoxia-stress adaptation in abalone.

3.5 Upcoming work

We are currently using transcriptome data available from *Haliotis* (J. Strugnell unpublished data) to annotate the outlier markers detected. We will use this information to identify genes that are associated with the local adaptation of these two species in different environments. We expect to finish analyses by the end of May 2015. We are planning to submit a manuscript entitled "Seascape genomics of Roe's abalone: a spatial analysis of neutral and adaptive variation in a species challenged by climate change" to a high impact journal by July 2015 and a second manuscript about Greenlip seascape genomics by August 2015. A minimum of another two manuscripts reporting on the extensive genomic datasets generated during this project and on implications for restocking and enhancement programs will also be prepared and submitted for publication in 2015.

3.6 Management implications of the results

3.6.1 Roe's abalone

- The original *H. roei* re-stocking attempts at Kalbarri Cliffs have successfully translocated adult Roe's abalone to the study sites at effective population sizes. Unfortunately juvenile (0+ age) abalone have not yet been detected and so the emphasis of the projects genetic component was changed to assessing and comparing the genomic diversity and differentiation between the remnant Kalbarri Cliff population and other *H. roei* populations around the Australian coast. Our advice now incorporates identifying which source populations would be suitable to use for restocking the Kalbarri Cliffs area.

- Levels of genome-wide diversity in *H. roei* are comparable to those of *H. laevigata* generated in our study. Levels of genome-wide diversity were similar in all populations analysed. The highest diversity was detected in the populations close to Perth (Roe's abalone).
    - Emphasis should none-the-less be placed on maximizing the genetic diversity of re-stocking efforts by using high founder numbers (e.g. to give genetically effective population sizes, N_e, greater than 500).
- The level of genetic diversity within the remnant Kalbarri Cliff population is comparable to that found within other *H. roei* populations around the coast (e.g. slightly higher than genetic diversity found within the neighbouring Lucky Bay population).
    - Because the sustained period of elevated water temperatures was a relatively recent event, there has not been time for loss of genetic diversity to occur from the remnant population. However, due to the low census and small effective population size of the remnant population we expect that loss of genetic variation would occur very rapidly in the next generations unless there is significant natural recruitment of abalone from neighbouring populations or extensive re-stocking programs are implemented to enhance recruitment.
- A minimum of three genetically distinct groups can be clearly defined in our dataset when we consider the section of the genome under selection (i.e. functionally important section of the genome). These are: 1. Kalbarri Cliff to Lucky Bay, 2. Greenough to Cape Augusta, and, 3. Albany to Spencer Gulf populations. Natural selection across the range has contributed significantly to the overall pattern of differentiation detected. Levels of genetic differentiation are generally low but highly significant between populations.
- If we consider only the fraction of the genomic dataset that is **not** under natural selection, the resulting pattern is of high connectivity and no to low population differentiation across the wide region sampled (i.e. from Kalbarri Cliff to Spencer Gulf). This is similar to patterns of connectivity reported for Australian abalones based on traditional genetic methods [e.g. 1] that are unable to pick up the signal of selection.
- The seascape analyses support the existence of three environmentally different regions, which are mostly congruent with the three adaptive clusters found in *H. roei*. The adaptive differentiation between the three genetic clusters of Roe's abalone is significantly correlated with a thermo geographic gradient.
    - Management should focus primarily on sourcing animals from the close geographical location of Lucky Bay (if possible) as this location shares similar environmental conditions to the collapsed population and as these individuals are more likely to contain useful adaptations promoting survival in the re-stocked environment. If *H. roei* numbers are low in Lucky Bay, and it is therefore not possible/desirable to source animals from this area, then sourcing animals from other areas further down the coast needs to be considered.

3.6.2 Greenlip abalone

- Levels of genome-wide diversity were similar in all populations analysed. The diversity was comparable to the diversity found in *H. roei* (this report). The highest diversity was detected in the easternmost sampled populations.
    - The use of large numbers of broodstock is recommended to maintain the genetic variability within and between populations.
- The "neutral" SNP dataset shows a pattern of high connectivity. This is similar to the pattern reported for Roe's abalone.
- When only the section of the genome under selection is considered, five genetically distinct groups can be clearly defined in our dataset. These are: 1. Outback to Windy Outside, 2. Parrys Bay and Whalebone Point, 3. Inner Island to Manson, 4. Fanny Cove to Burton Rocks, and 5. Rob Island to Gulch .
- The seascape analysis supports the existence of four oceanographically different regions, which are partially congruent with the five adaptive clusters in *H. laevigata*. The adaptive differentiation between genetic clusters is significantly correlated with geographic variation in oxygen concentration.
    - The stock enhancing program should use broodstock taken directly from the genetic group to be enhanced. If broodstock cannot be taken from the genetic group to be enhanced, they should be taken from the genetically most similar group available. Since the differentiation detected is adaptive, the stock enhancing program should be genetically monitored to detect the contribution of the broodstocks to the next generation.

In summary, we have made the following progress against objectives:

1. *To develop new expertise and transfer genomic approaches (i.e. GBS or RAD-seq), currently used for commercially important species in the northern Hemisphere, to benefit the fisheries or aquaculture of key commercial species in Australia*. This objective has been fully achieved. The post-doc has successfully adapted and applied these approaches to investigate stock structure of the species in question.
2. *To generate and use genomic data to analyse changes and patterns in functionally relevant genetic diversity (i.e. variation related to fitness) with restocking and ranching and advise on management strategies for the species concerned*. This objective has been achieved. Advice on management strategies for the species is being communicated to the Department of Fisheries Western Australia based on the stock structure results. The post-doc is currently determining whether the detected "functional variation" (ecologically relevant variation related to the fitness of individuals) is associated with oceanographic data to test for correlations

between environmental and genetic factors. The post-doc is also attempting to annotate the outlier markers detected to determine what genes are involved.

3. *Through objective 2, to create genetically healthy/sustainable restocked populations*. This objective has been partly met (see above).
4. *Through objective 2, to improve returns from ranching while having minimal impact on natural populations*. This objective has been partly met (see above).
5. *Through objective 2, to maintain healthy fisheries*. This objective has been partly met (see above).
6. *To develop expertise in genetic management associated with enhancement and related interventions*. This objective has been partly met. The project team has met and had preliminary discussions regarding genetic management and its application to enhancement and related interventions. The project team will meet again to fully discuss the implications of the results.

5. Benefits and Adoption

Restocking is playing an important role in the conservation and management of marine species under exploitation. Considering the global per capita marine product consumption in the last decades, the restoration of populations that have been seriously depleted will help to secure the production from fisheries with important social and economic benefits. The restocking of Roe's abalone in the Kalbarri region would deliver considerable benefits, particularly for the local recreational fisheries; it will create a source of seafood and income for the local communities.

The genomic resources we have produced will help to address fundamental questions for the efficient restocking of Roe's and stock enhancement of Greenlip abalone in Western Australia. We have identified three possible adaptive stocks of the species, and have developed a catalogue of more than 30,000 SNPs for each species. These can be used as genetic markers for the constant monitoring of the wild populations and are important tools for the fisheries management of the species over its whole distribution.

Moreover, the distinction of outlier loci would benefit the aquaculture industry. By annotating these loci we can identify genes involved in physiological, morphological and behavioural traits; information that could be used to optimize individual selection and mating programs in order to increase productivity.

6. Further Development

Once new animals are recruited in the restocked population, the population of new recruits should be sampled and tested using the same RAD-seq technique. It should be possible to use these data to assess the success of contribution by particular parents and the genetically effective population size of the founding generations. Selective pressures at particular loci in the re-stocked population could also be monitored over time. This will be important information to use for guiding the management of the restocked population (i.e. should further restocking be attempted, where should the animals be stocked, what source of stock to use).

Additional samples should be genotyped across the range of Greenlip abalone to test for population structure, levels of genetic diversity, selective sweeps and genetic correlations with geospatial data. This information would be useful for making more informed fishery management decisions for the Greenlip populations in WA.

7. Planned Outcomes

7.1 Public Benefit Outcomes

- Improvement of the conservation and management of *H. roei* and *H. laevigata* stocks directly benefits local recreational and commercial fisheries, promoting stock sustainability.
- Sustainable stock will provide economic, social and ecological benefits.

7.2 Private Benefit Outcomes

- Genomic resources for the monitoring of *H. roei* and *H. laevigata* wild and hatchery stocks.
- Potential genomic resources for the improvement of mass selection and mating schemes in aquaculture activities.

7.3 Linkages with CRC Milestone Outcomes

- Comprehensive genomic analysis of population genetic diversity and population connectivity in *H. roei* stocks in Western and South Australia;
- Comprehensive genomic analysis of population genetic diversity and population connectivity in *H. laevigata* stocks in Western Australia;
- Recommendations for capturing genetic diversity and locally adapted genotypes for increasing the chances of re-stocking and stock enhancement success;
- Development of genomic profiling and next generation sequencing capacity for the Australian Seafood CRC.

8. Conclusions

The *H. roei* restocking at Kalbarri Cliffs has successfully translocated adult Roe's abalone to the study sites at “effective” population sizes.

Levels of genome-wide diversity were similarly moderate to high in all *H. roei* and *H. laevigata* populations analysed, including the *H. roei* population from Kalbarri Cliff.

The neutral fraction of the genomic dataset shows a pattern of isolation by distance with higher connectivity between proximate localities than between distant localities in both species.

For *H.roei*, three distinct groups can be clearly defined in our genomic dataset target by natural selection. These are: 1. Kalbarri Cliff to Lucky Bay, 2. Greenough to Augusta, and 3. Albany to Spencer Gulf populations.

For *H.laevigata*, five distinct groups can be clearly defined in our dataset targeted by natural selection. These are 1. Outback to Windy Outside, 2. Parrys Bay and Whalebone Port, 3. Inner Island to Mason, 4. Fanny Cove to Burton Rocks, and 5. Rob Island to Gulch (Fig 23).

The pattern of adaptive genetic differentiation appears correlated with a thermal geographic gradient in *H. roei* and with differences in oxygen concentration in *H. laevigata.*

The genotyping by sequencing protocols used in this work were very efficient for assessing genome-wide diversity in the two study species of abalone. These protocols can be implemented for other commercially important species using the capabilities developed during this project.

9. References

1 **Piggott M, Banks S, Tung P and Beheregaray LB: Genetic evidence for different scales of connectivity in a marine mollusc. *Marine Ecology Progress Series.* 2008, 365: 127-136.**

2 **Baird NA, DEtter PD, Atwood TS, Currey MC and Shiver AL: Rapid SNP discovery and genetic mapping using sequenced RAD markers. *PLoS One.* 2008, 3: e3376.**

3 **Peterson BK, Weber JN, Kay EH, Fisher HS and Hoekstra HE: Double digest RADseq: An inexpensive method for *de novo* SNP discovery and genotyping in modela and non-model species. *PLoS One.* 2012, 7(5): e37135.**

4 **Altshuler D, Pollara VJ, Cowles CR, Van Etten WJ, Baldwin J, Linton L and Lander ES: An SNP map of the human genome generated by reduced representation shotgun sequencing. *Nature.* 2000, 407(6803): 513 - 516.**

5 **Xu P, Xu SZ, Wu XH, Tao Y, Wang BG, Wang S, Qin DH, Lu ZF and Li GJ: Population genomic analyses from low-coverage RAD-Seq data: a case study on the non-model cucurbit bottle gourd. *Plant Journal.* 2014, 77(3): 430-442.**

6 **Richards PM, Liu MM, Lowe N, Davey JW, Blaxter ML and Davison A: RAD-Seq derived markers flank the shell colour and banding loci of the *Cepaea nemoralis* supergene. *Molecular Ecology.* 2013, 22(11): 3077-3089.**

7 **Cariou M, Duret L and Charlat S: Is RAD-seq suitable for phylogenetic inference? An in silico assessment and optimization. *Ecology and Evolution.* 2013, 3(4): 846-852.**

8 **Gagnaire PA, Pavey SA, Normandeau E and Bernatchez L: The genetic architecture of reproductive isolation during speciation-with-gene-flow in lake whitefish species pairs assessed by RAD sequencing. *Evolution.* 2013, 67(9): 2483-2497.**

9 **Davey JL and Blaxter MW: RADSeq: next-generation population genetics. *Briefing in Functional Genomics.* 2010, 9: 416-423.**

10 **Peterson BK, Weber JN, Kay EH, Fisher HS and Hoekstra HE: Double digest RADseq: an inexpensive method for de novo SNP discovery and genotyping in model and non-model species. *PloS one.* 2012, 7(5): e37135.**

11 **Catchen JM, Amores A, Hohenlohe P, Cresko W and Postlethwait JH: Stacks: building and genotyping loci de novo from short-read sequences. *Genes, Genomes, Genetics.* 2011, 1: 171-187.**

12 **Excoffier L and Lischer HEL: Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resources.* 2010, 10(3): 564-567.**

13 **Pritchard J, Stephens M and Donnelly J: Inference of population structure using multilocus genotype data. *Genetics.* 2000, 155: 945-959.**

14 **Jombart T and Ahmed I: adegenet 1.3-1: new tools for the analysis of genome-wide SNP data. *Bioinformatics.* 2011, 27(21): 3070-3071.**

15 **Meirmans PG and Van Tienderen PH: GENOTYPE and GENODIVE: two programs for the analysis of genetic diversity of asexual organisms. *Molecular Ecology Notes.* 2004, 4(4): 792-794.**

16 **Hart AM, Fabris F, Brown J and Caputi N: Biology, history, and assessment of Western Australian abalone fisheries. *Fisheries Research Report, Department of Fisheries, Western Australia.* 2013, 90.**

17 **Schlitzer R (2010) Ocean Data View.**

18 **Lê S, Josse J and Husson F: FactoMineR: an R package for multivariate analysis. *Journal of statistical software.* 2008, 25(1): 1-18.**

19 **Goslee SC and Urban DL: The ecodist package for dissimilarity-based analysis of ecological data. *Journal of Statistical Software.* 2007, 22(7): 1-19.**

20 **Dixon P: VEGAN, a package of R functions for community ecology. *Journal of Vegetation Science.* 2003, 14(6): 927-930.**

21 **Gruenthal K, Witting D, Ford T, Neuman M, Williams J, Pondella II D, Bird A, Caruso N, Hyde J, Seeb L and Larson W: Development and application of genomic tools to the restoration of green abalone in southern California. *Conservation Genetics.* 2013: DOI 10.1007/s10592-10013-10524-10595.**

10. Appendix

Appendix 1. Estimated annual average of the fiver oceanographic variables and maximum sea surface temperature during the the 2010/2011 marine heatwave event (SSTHW).

ID	pH	Nutrients Concentration	Oxygen Concentration	Sea Surface Temperature	Maximum Sea Surface Temperature	Maximum SSTHW
Kalbarri Cliff	8.26	2.62	4.28	22.33	24.47	28.44
Lucky Bay	8.26	2.61	4.38	22.25	24.23	28.02
Greenough	8.25	2.31	4.98	20.64	23.48	27.37
Burns Beach	8.20	2.13	5.28	20.60	22.05	25.65
Waterman	8.20	2.18	5.26	20.59	22.02	25.54
Penguin Island	8.19	2.15	5.20	20.65	22.37	25.20
Outback Middle	8.25	3.17	5.21	20.01	21.55	23.95
Augusta	8.25	3.23	5.19	19.89	21.50	23.90
Coral Patch	8.25	3.37	5.17	19.92	21.49	23.88
Windy Outside	8.25	3.27	5.20	19.47	20.66	23.05
Albany	8.28	2.97	5.14	19.74	20.32	22.20
Parrys Bay	8.24	2.81	5.14	19.76	20.32	21.94
Inner Island	8.24	3.06	5.15	19.40	20.32	21.04
Whalebone Pt	8.22	2.59	5.12	18.92	21.03	20.66
2 Mile Primary	8.26	2.66	5.47	18.79	21.08	20.66
Masons	8.28	2.71	5.44	18.53	20.88	20.60
Fanny Cove	8.28	2.64	5.38	18.58	20.52	20.57
Burton Rocks	8.28	2.77	5.38	18.63	20.38	20.57
Rob Island	8.32	3.01	5.37	18.69	20.29	20.45
Ben Island	8.31	2.91	5.38	18.79	20.48	20.48
Esperance	8.32	3.29	5.39	18.68	20.59	20.41
Gulch	8.32	3.03	5.39	18.69	20.59	20.40
Spencer Gulf	8.15	1.49	5.24	19.67	20.45	18.19